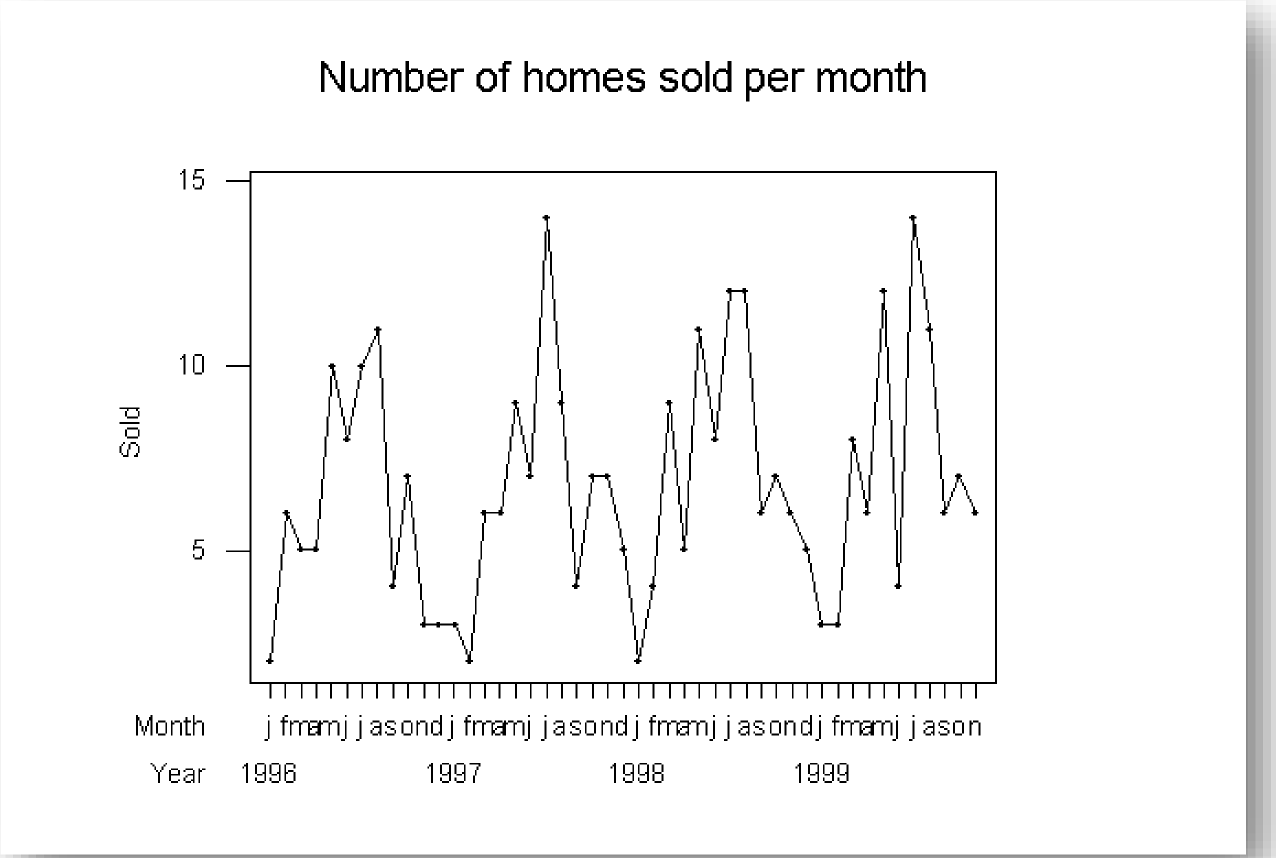
Föreläsning 2
Klassisk komponent-uppdelning

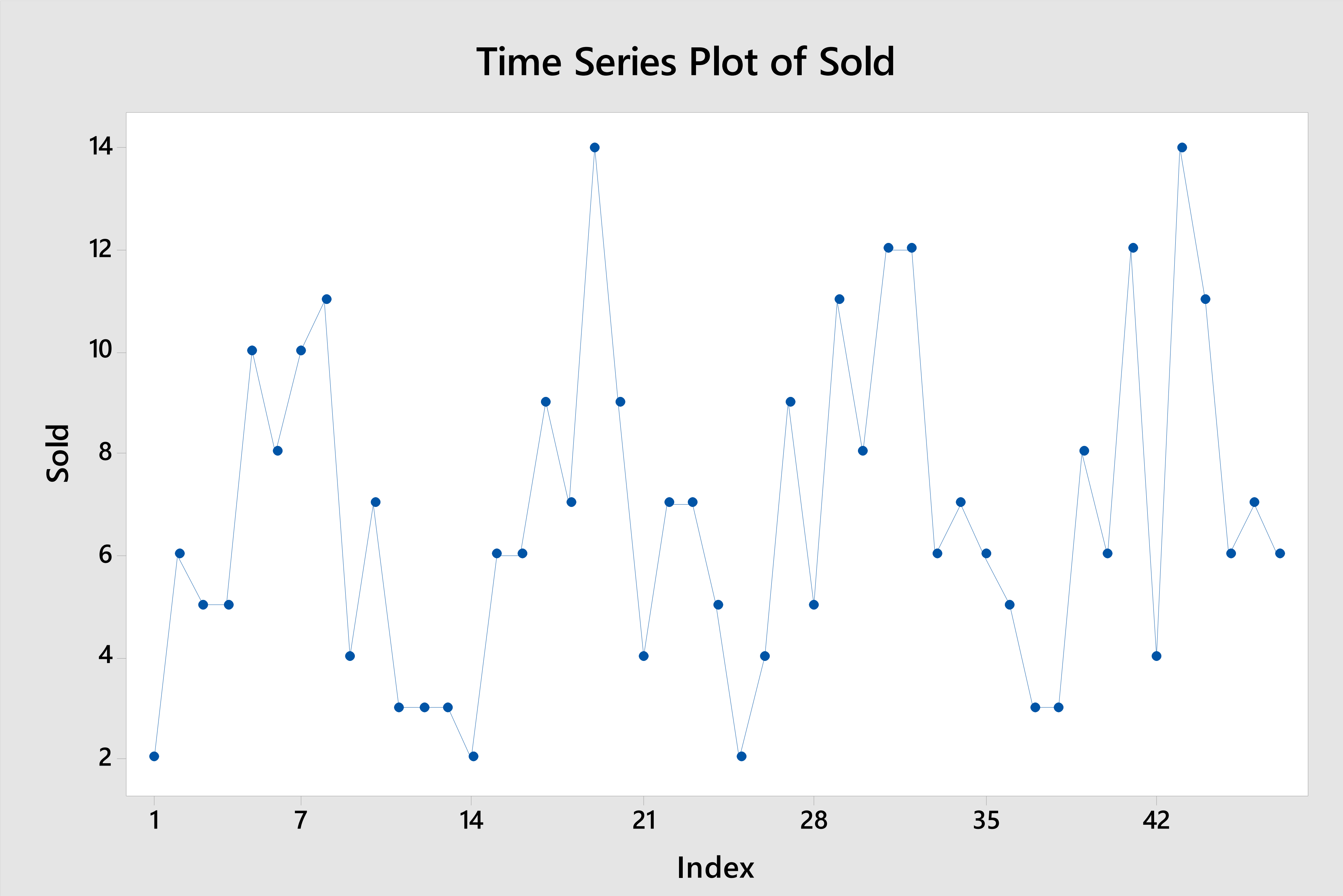
Trend- och säsongsjustering

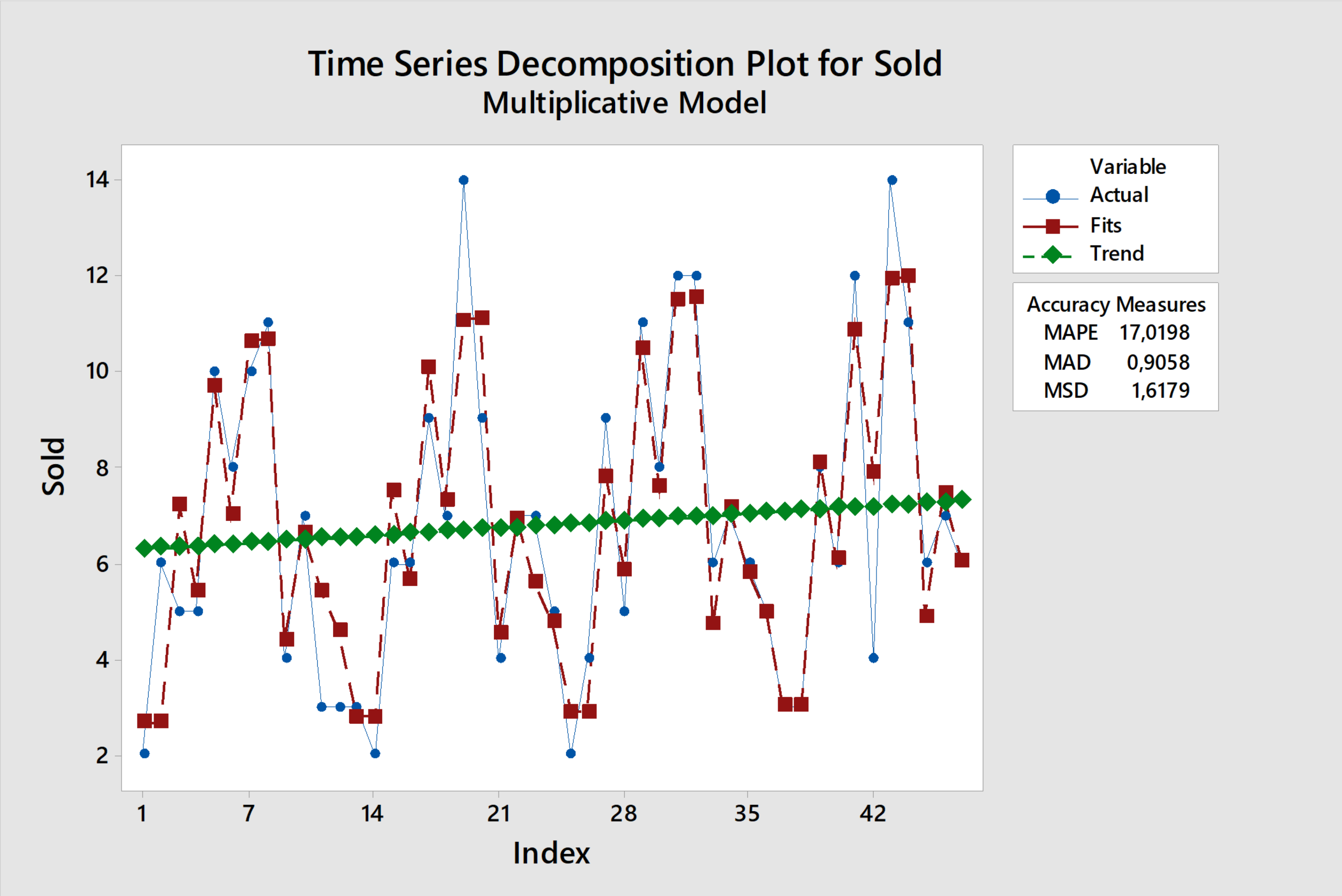
Kapitel 2,4,2

Åter till ex med Sold, antal sålda hus över tiden.

- Typ av trend
- Typ av säsongsvariation
- Additivt eller multiplikativt mönster
- Statiskt eller dynamiskt mönster
- Tolka serien eller beräkna prognos




Time Series Plot of Sold
Sold
14
12
10
8
6
4
2
1
7
14
21
28
35
42
Index



Time Series Decomposition Plot for Sold
Multiplicative Model
Sold
Index
Variable
Actual
Fits
Trend
Accuracy Measures
MAPE 17,0198
MAD 0,9058
MSD 1,6179
14
12
10
8
6
4
2
1
7
14
21
28
35
42

Denna metod fungerar bra om tidsserien uppvisar ett statiskt mönster.

Det är fyra komponenter i modellen:

- Multiplikativ modell:

$$y_t = TR_t \cdot SN_t \cdot CL_t \cdot IR_t$$

- Additiv modell:

$$y_t = TR_t + SN_t + CL_t + IR_t$$

där

- TR_t = Trendkomponent
- SN_t = Säsongkomponent
- CL_t = Cykliska komponent
- IR_t = Slumpkomponent

Beteckna den skattade trendkomponenten TR med tr,

skattad säsongskomponent SN med sn

skattad cyklisk komponent CL med cl (Denna ska vi inte inkludera)

samt den skattade slumpkomponenten IR med ir.

Idén är nu att vi ska skatta en komponent i taget och därefter rensa bort den. Detta görs tills det endast finns slump kvar dvs IR.

Detta görs helt annorlunda jämfört med tidsserieregression där alla komponenter skattas samtidigt.

En metod då alla parametrar/komponenter skattas samtidigt är att föredra!!

Skattning och rensning av komponenter

Säsongsrensning: Borttagandet av säsongsvariation

y_t - sn_t i den additiva modellen

y_t / sn_t i den multiplikativa modellen

Säsongsvariation överskuggar ofta andra relevanta komponenter. Genom säsongrensningen kan man alltså enklare se trender och andra komponenter.

Allmänt om glidande medelvärden beskrivs på tavlan.

Skattning av säsongskomponenten samt säsongsrensning

Säsongrensning:

Serien *rensas* från säsongkomponenten genom beräkning av *centrerade och viktade glidande medelvärden (centered moving averages, CMA)*:

$$CMA_t = \frac{y_{t-(L/2)} + 2y_{t-(L/2-1)} + \ldots + 2y_t + \ldots + 2y_{t+(L/2-1)} + y_{t+(L/2)}}{2L}$$

där L = Säsongslängden i serien. Ex: L=2 för halvårsdata, 4 för kvartalsdata och 12 för månadsdata. OBS Formeln ovan gäller när L är ett jämnt tal.

Exempel (sold data från tidigare) Multiplikativ modell

tid	månad	antal	CMA
1	1	2	*
2	2	6	*
3	3	5	*
4	4	5	*
5	5	10	*
6	6	8	*
7	7	10	6.21
8	8	11	6.08
9	9	4	5.95
10	10	7	....
11	11	3	
12	12	3	
13	1	3	
14	2	2	
15	3	6	

$$
\begin{aligned}
&(2+ \\
&+2\cdot 6+2\cdot 5+2\cdot 5+2\cdot 10+2\cdot 8+2\cdot 10+ \\
&+2\cdot 11+2\cdot 4+2\cdot 7+2\cdot 3+2\cdot 3+ \\
&+3)/(2\cdot 12)
\end{aligned}
$$

En första skattning av så kallade grova säsongkomponenter erhålls genom att beräkna

- y_t/CMA_t i en multiplikativ modell
- $y_t - CMA_t$ i en additiv modell

och sen beräkna medelvärden för alla värden som avser samma säsong. (t.ex. alla januari-värden av y_t/CMA_t, etc.) → Totalt *L* medelvärden.

Tid	Mån.	Sold	CMA	Grova säs.kom.	Tid	Mån.	Sold	CMA	Grova säs.kom.
1	1	2	*	*	25	1	2	7.00000	**0.28571**
2	2	6	*	*	26	2	4	7.04167	**0.56805**
3	3	5	*	*	27	3	9	7.25000	**1.24138**
4	4	5	*	*	28	4	5	7.33333	**0.68182**
5	5	10	*	*	29	5	11	7.29167	**1.50857**
6	6	8	*	*	30	6	8	7.25000	**1.10345**
7	7	10	6.20833	**1.61074**	31	7	12	7.29167	**1.64571**
8	8	11	6.08333	**1.80822**	32	8	12	7.29167	**1.64571**
9	9	4	5.95833	**0.67133**	33	9	6	7.20833	**0.83237**
10	10	7	6.04167	**1.15862**	34	10	7	7.20833	**0.97110**
11	11	3	6.04167	**0.49655**	35	11	6	7.29167	**0.82286**
12	12	3	5.95833	**0.50350**	36	12	5	7.16667	**0.69767**
13	1	3	6.08333	**0.49315**	37	1	3	7.08333	**0.42353**
14	2	2	6.16667	**0.32432**	38	2	3	7.12500	**0.42105**
15	3	6	6.08333	**0.98630**	39	3	8	7.08333	**1.12941**
16	4	6	6.08333	**0.98630**	40	4	6	7.08333	**0.84706**
17	5	9	6.25000	**1.44000**	41	5	12	*	*
18	6	7	6.50000	**1.07692**	42	6	4	*	*
19	7	14	6.54167	**2.14013**	43	7	14	*	*
20	8	9	6.58333	**1.36709**	44	8	11	*	*
21	9	4	6.79167	**0.58896**	45	9	6	*	*
22	10	7	6.87500	**1.01818**	46	10	7	*	*
23	11	7	6.91667	**1.01205**	47	11	6	*	*
24	12	5	7.04167	**0.71006**					

Medelvärden av grova säsongskomponenter:

Juli: $(1.61074+2.14013+1.64571)/3 \approx 1.7989$

Aug: $(1.80822+1.36709+1.64571)/3 \approx 1.6070$

Sep: $(0.67133+0.58896+0.83237)/3 \approx 0.6976$

Okt: $(1.15862+1.01818+0.97110)/3 \approx 1.0493$

Nov: $(0.49655+1.01205+0.82286)/3 \approx 0.7772$

Dec: $(0.50350+0.71006+0.69767)/3 \approx 0.6371$

Jan: $(0.49315+0.28571+0.42353)/3 \approx 0.4008$

Feb: $(0.32432+0.56805+0.42105)/3 \approx 0.4378$

Mar: $(0.98630+1.24138+1.12941)/3 \approx 1.1190$

Apr: $(0.98630+0.68182+0.84706)/3 \approx 0.8384$

Maj: $(1.44000+1.50857)/2 \approx 1.4743$ *Obs! Bara två värden här!*

Juni: $(1.07692+1.10345)/2 \approx 1.0902$ *…och här!*

Medelvärdena måste sedan justeras så att de

- vid multiplikativ modell får medelvärde 1*, (*dvs summan av alla justerade säsongmedelvärden ska bli *L)*

- vid additiv modell fås medelvärde 0, (dvs summan av alla justerade säsongmedelvärden ska bli 0.)

De justerade värdena kallas för säsongskomponenter

$sn_1, \ldots, sn_L$

Summan av de beräknade medelvärdena:

$$1.7989 + 1.6070 + 0.6976 + 1.0493 + 0.7772 + 0.6371 + 0.4008 + 0.4378 + 1.1190 + 0.8384 + 1.4743 + 1.0902 \approx 11.9276$$

Summan ska bli L=12

För att få den till 12 multipliceras samtliga medelvärden med

$$12/11.9276 \approx 1.00607$$

Slutligt skattade säsongkomponenter:

Jan: $sn_1 = 0.4008 \cdot 1.00607 \approx 0.403$

Feb: $sn_2 = 0.4378 \cdot 1.00607 \approx 0.440$

Mar: $sn_3 = 1.1190 \cdot 1.00607 \approx 1.126$

Apr: $sn_4 = 0.8384 \cdot 1.00607 \approx 0.843$

Maj: $sn_5 = 1.4743 \cdot 1.00607 \approx 1.483$

Juni: $sn_6 = 1.0902 \cdot 1.00607 \approx 1.097$

Juli: $sn_7 = 1.7989 \cdot 1.00607 \approx 1.809$

Aug: $sn_8 = 1.6070 \cdot 1.00607 \approx 1.617$

Sep: $sn_9 = 0.6976 \cdot 1.00607 \approx 0.702$

Okt: $sn_{10} = 1.0493 \cdot 1.00607 \approx 1.056$

Nov: $sn_{11} = 0.7772 \cdot 1.00607 \approx 0.782$

Dec: $sn_{12} = 0.6371 \cdot 1.00607 \approx 0.641$

Tidsserien säsongrensas genom:

- $d_t = y_t / sn_t$ vid multiplikativ modell

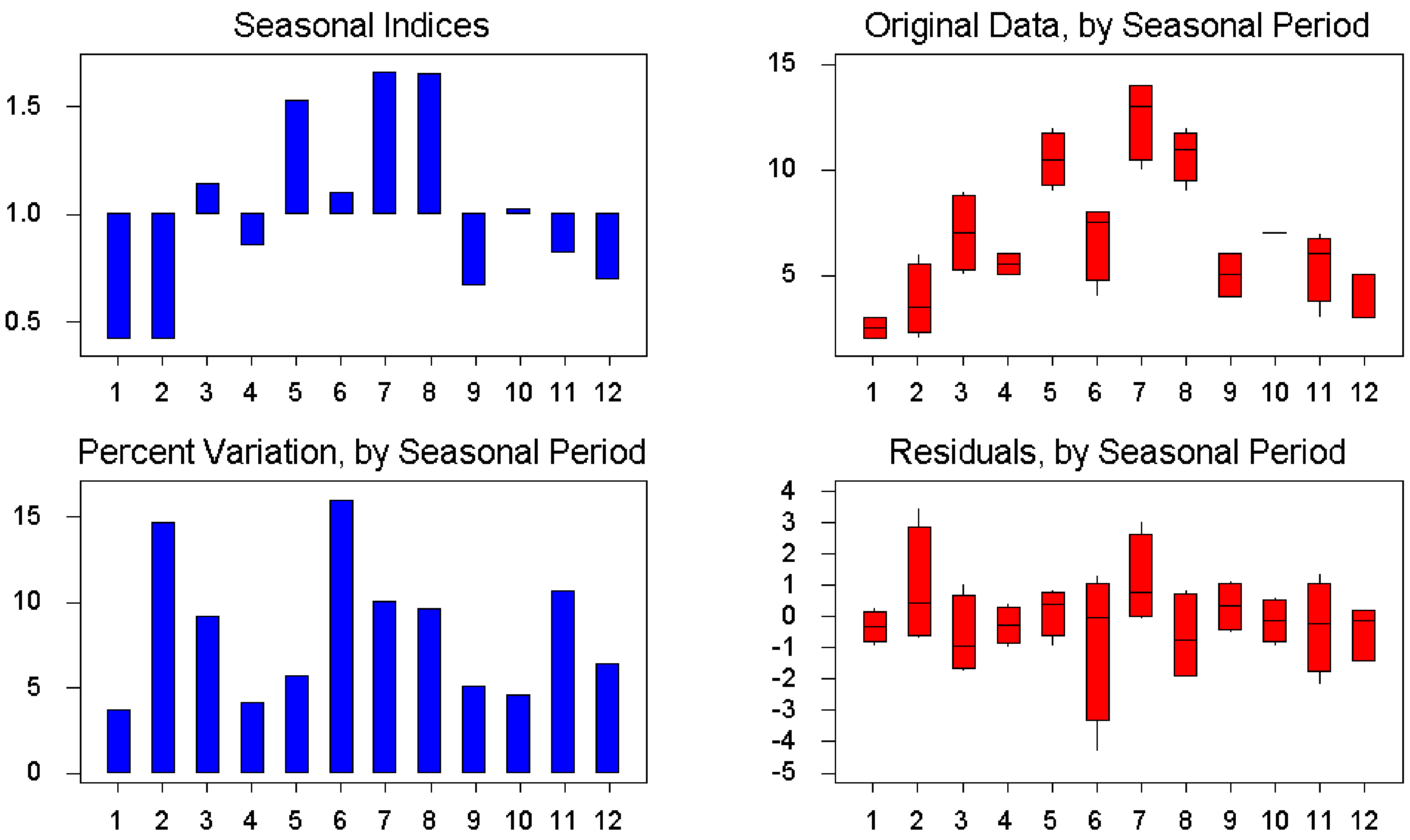
- $d_t = y_t - sn_t$ vid additiv modell

där sn_t är något av värdena $sn_1, \ldots, sn_L$

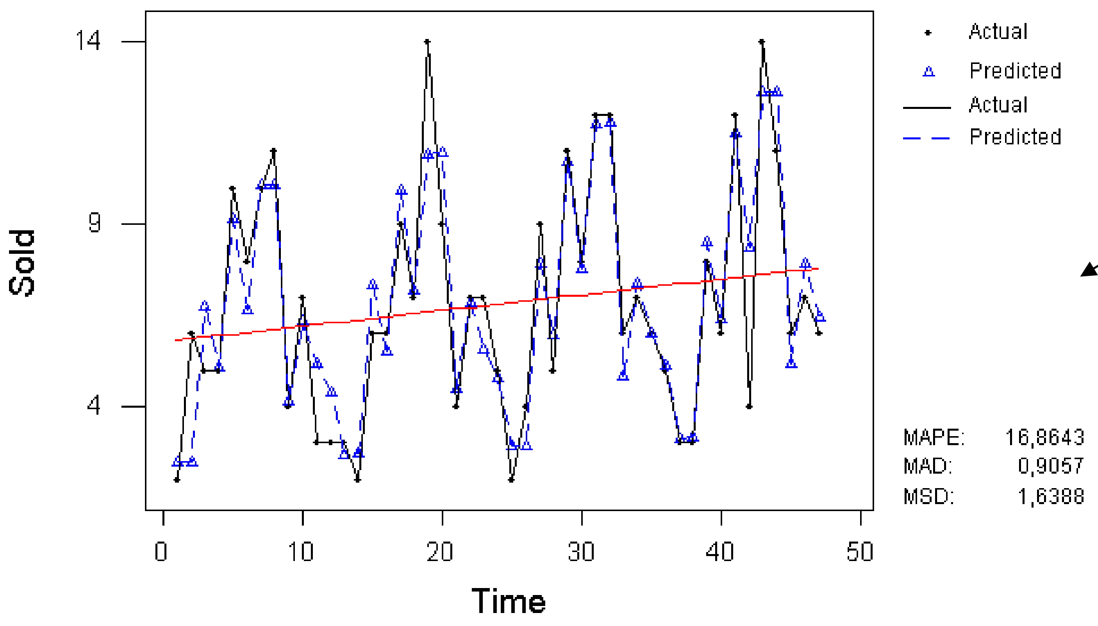
beroende på vilken av säsongerna som *t* motsvarar.

Tid	Mån.	Sold y_t	sn_t	d_t	Tid	Mån.	Sold y_t	sn_t	d_t
1	1	2	0.403	4.963	25	1	2	0.403	4.963
2	2	6	0.44	13.636	26	2	4	0.44	9.091
3	3	5	1.126	4.440	27	3	9	1.126	7.993
4	4	5	0.843	5.931	28	4	5	0.843	5.931
5	5	10	1.483	6.743	29	5	11	1.483	7.417
6	6	8	1.097	7.293	30	6	8	1.097	7.293
7	7	10	1.809	5.528	31	7	12	1.809	6.633
8	8	11	1.617	6.803	32	8	12	1.617	7.421
9	9	4	0.702	5.698	33	9	6	0.702	8.547
10	10	7	1.056	6.629	34	10	7	1.056	6.629
11	11	3	0.782	3.836	35	11	6	0.782	7.673
12	12	3	0.641	4.680	36	12	5	0.641	7.800
13	1	3	0.403	7.444	37	1	3	0.403	7.444
14	2	2	0.44	4.545	38	2	3	0.44	6.818
15	3	6	1.126	5.329	39	3	8	1.126	7.105
16	4	6	0.843	7.117	40	4	6	0.843	7.117
17	5	9	1.483	6.069	41	5	12	1.483	8.092
18	6	7	1.097	6.381	42	6	4	1.097	3.646
19	7	14	1.809	7.739	43	7	14	1.809	7.739
20	8	9	1.617	5.566	44	8	11	1.617	6.803
21	9	4	0.702	5.698	45	9	6	0.702	8.547
22	10	7	1.056	6.629	46	10	7	1.056	6.629
23	11	7	0.782	8.951	47	11	6	0.782	7.673
24	12	5	0.641	7.800					

Seasonal Analysis for Sold



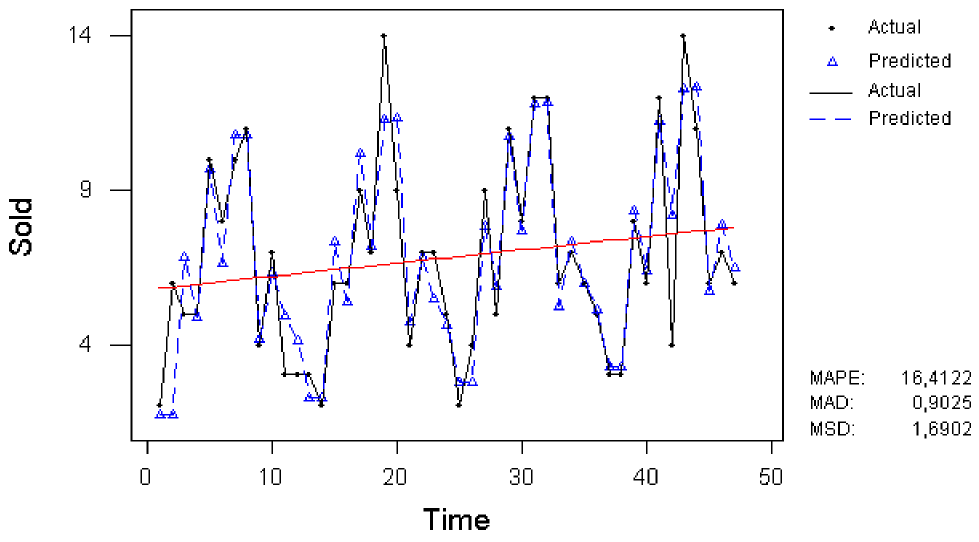
Decomposition Fit for Sold



Multiplikativ

Additiv

Decomposition Fit for Sold



Därefter skattas trendkomponenten Trendanalys

- Oavsett om tidsserien har säsongsvariation eller annan variation så kan vi göra en analys av trenden. Trendfunktonen betecknas TR

Data med endast linjär trend:

$$y_t = TR_t + \varepsilon_t = \beta_0 + \beta_1 \cdot t + \varepsilon_t$$

Data med kvadratisk trend:

$$y_t = TR_t + \varepsilon_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \varepsilon_t$$

Data med kubisk trend:

$$y_t = TR_t + \varepsilon_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + \varepsilon_t$$

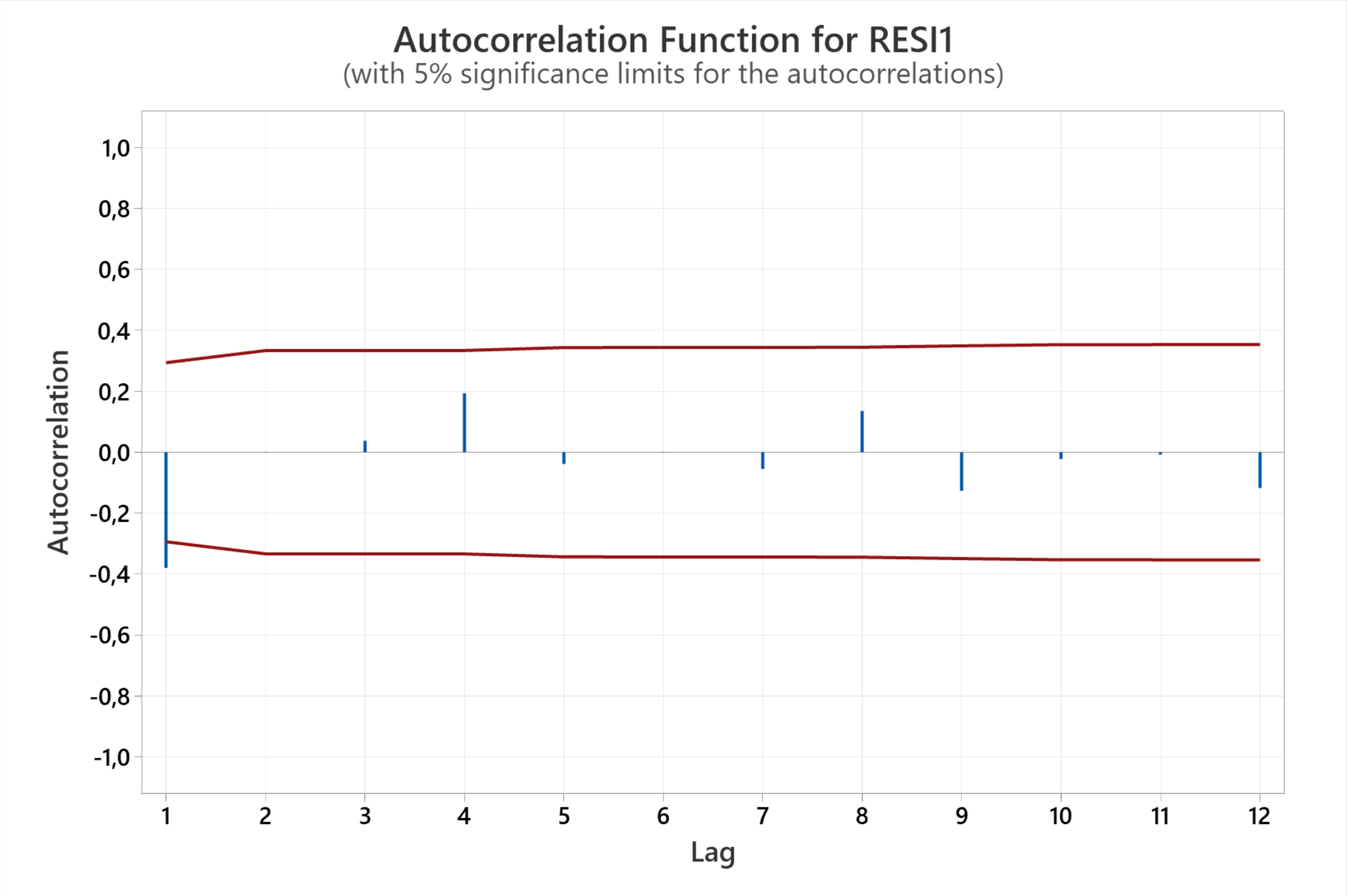
```
Trend Line Equation

Yt = 5.77613 + 4.30E-02*t

Seasonal Indices

Period     Index

   1       0.425997
   2       0.425278
   3        1.14238
   4       0.856404
   5        1.52471
   6        1.10138
   7        1.65646
   8        1.65053
   9       0.670985
  10        1.02048
  11       0.825072
  12       0.700325
```

← multiplikativ

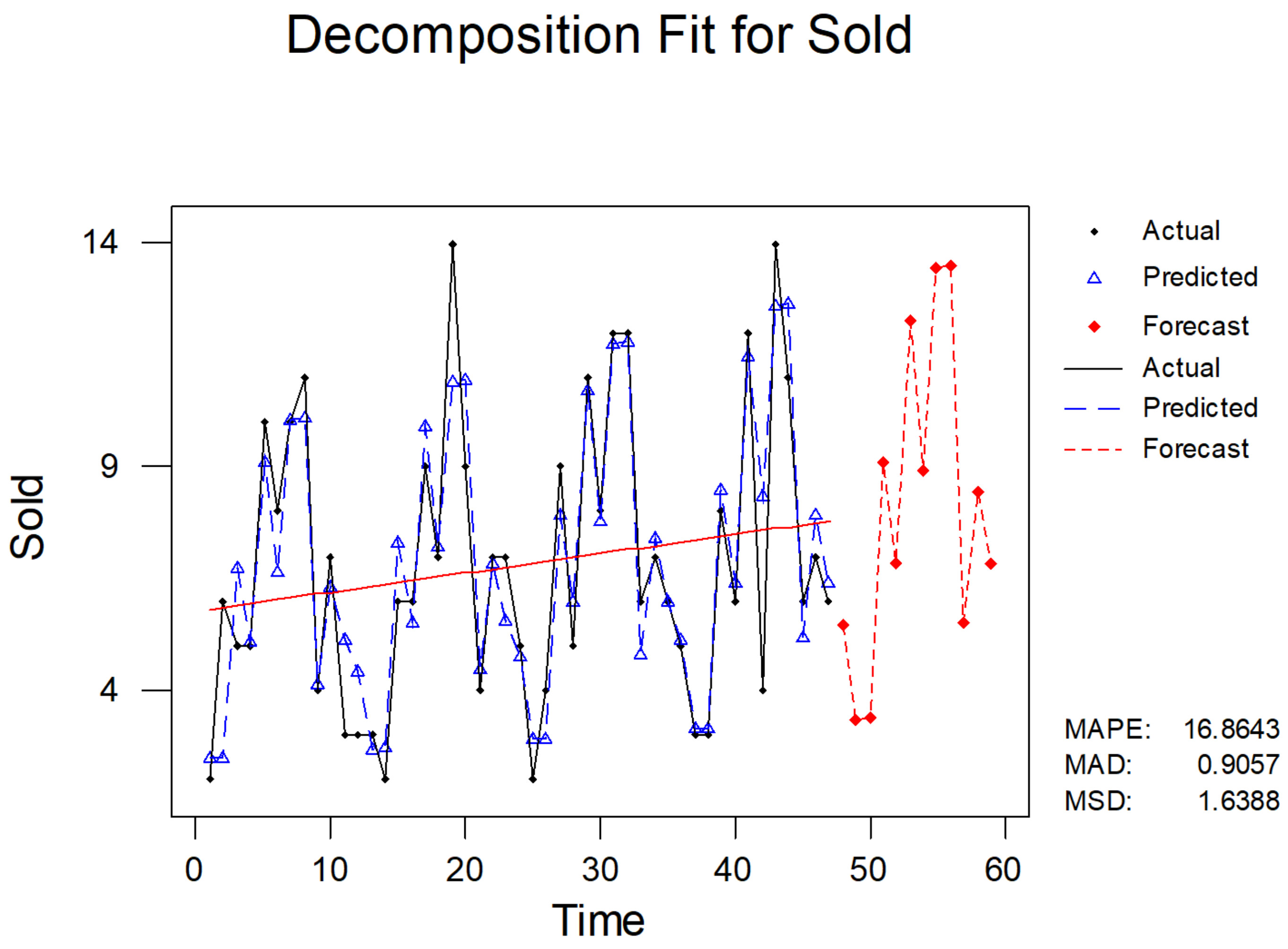
additiv →

```
Trend Line Equation

Yt = 5.77613 + 4.30E-02*t

 Seasonal Indices

Period     Index

   1       -4.09028
   2       -4.13194
   3       0.909722
   4       -1.09028
   5        3.70139
   6       0.618056
   7        4.70139
   8        4.70139
   9       -1.96528
  10       0.118056
  11       -1.29861
  12       -2.17361
```

SAC på residualer Additiv modell



Prognoser i statiska modeller



Prognoser i en klassisk komponentuppdelningsmodell

```
Trend Line Equation

Yt = 5.77613 + 4.30E-02*t


Seasonal Indices

Period     Index

   1       0.425997
   2       0.425278
   3        1.14238
   4       0.856404
   5        1.52471
   6        1.10138
   7        1.65646
   8        1.65053
   9       0.670985
  10        1.02048
  11       0.825072
  12       0.700325
```

Prognos för december 1999

tid t=48, säsong=1

multiplikativ modell

$$\hat{y} = (5.776 + 0.043 \cdot 48) \cdot 0.7 = 5.49$$

Prognos för januari 2000:

$$\hat{y} = (5.776 + 0.043 \cdot 49) \cdot 0.426 = 3.36$$

Detta var genomgång av:

- Kapital 2,4,2