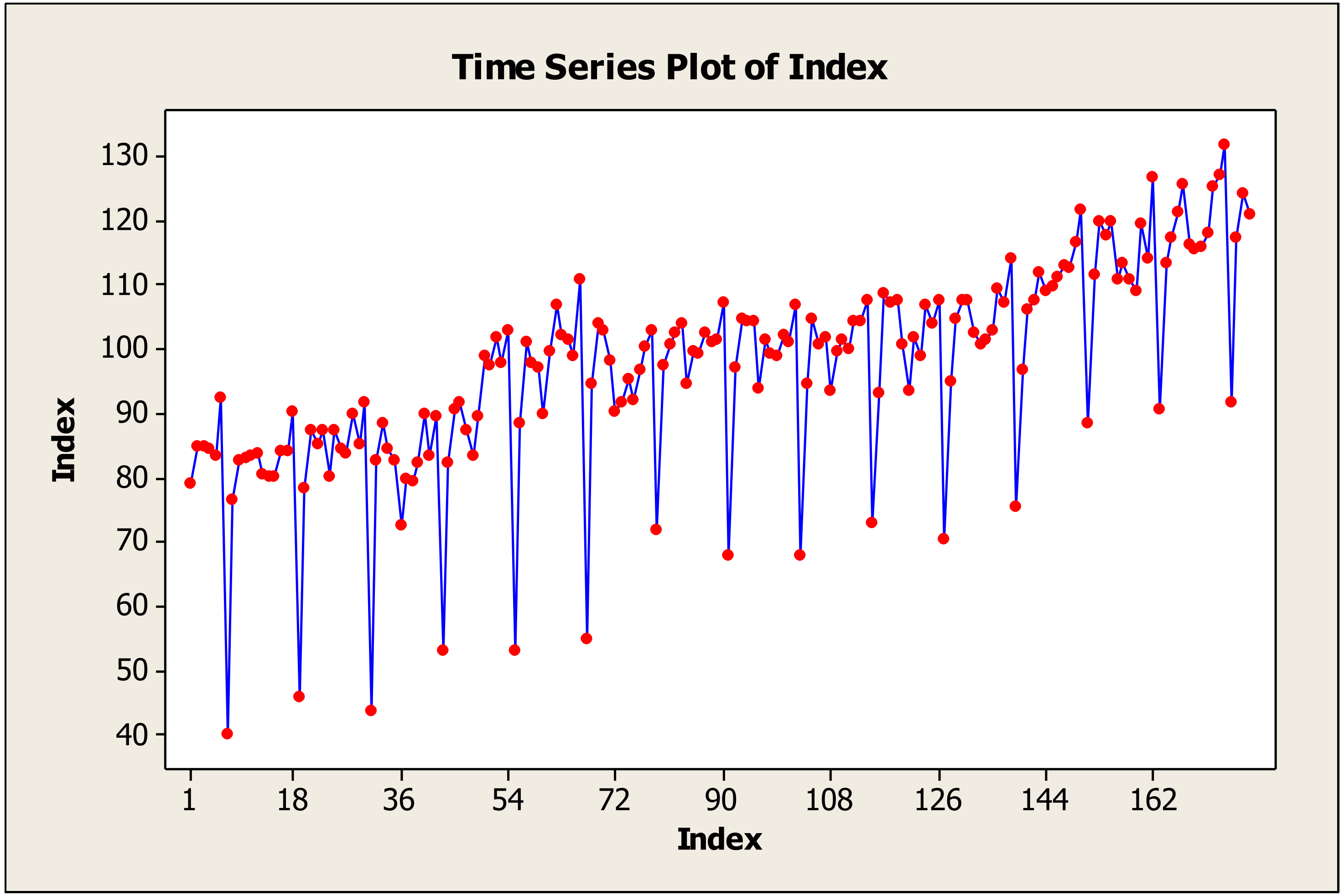
Föreläsning 7

Fortsättning ARIMA och SäsongsARIMA Prognoser för ARIMA-modeller och SARIMA-modeller

Forts Kap 5

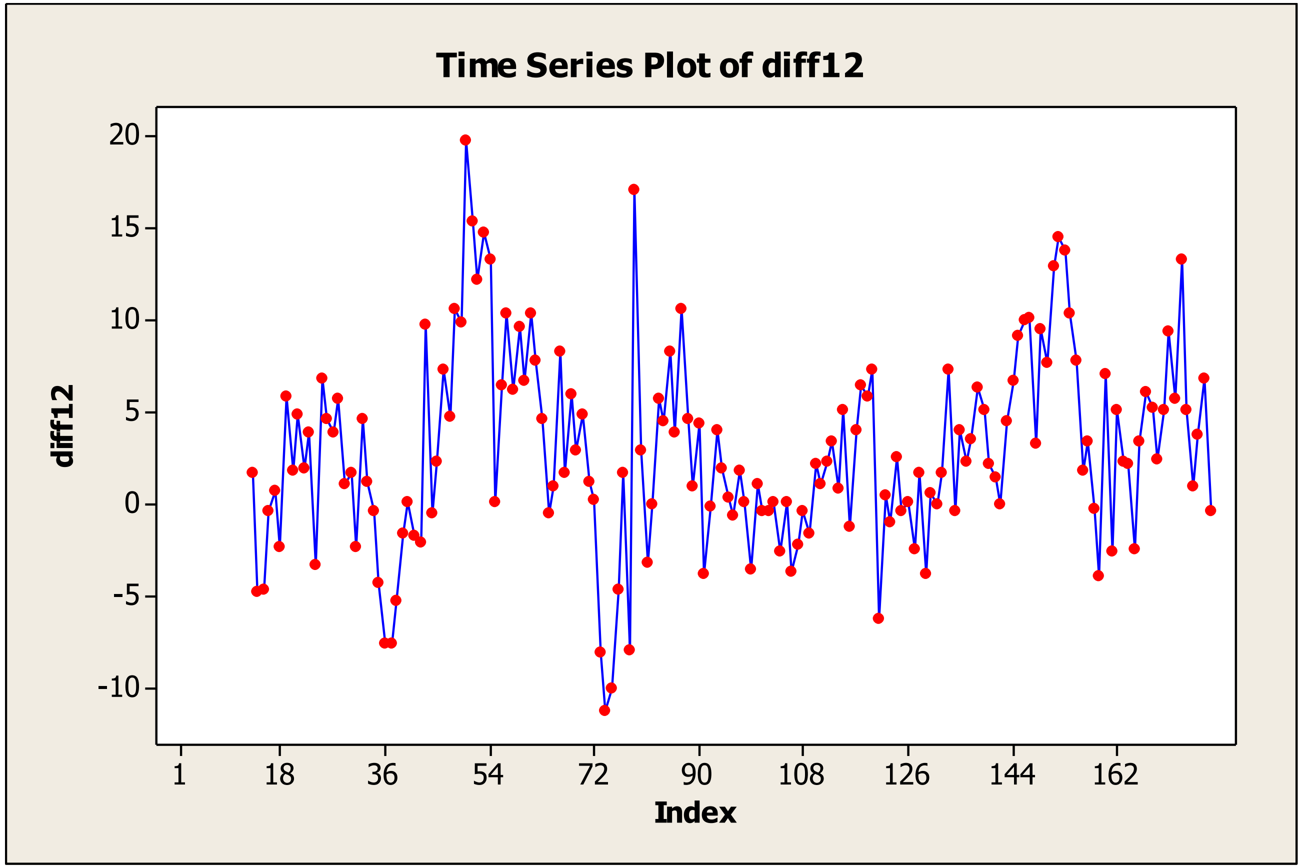
Forts Pappersproduktion



Vi har följt punkt 1. Vi fortsätter nu med punkt 2.

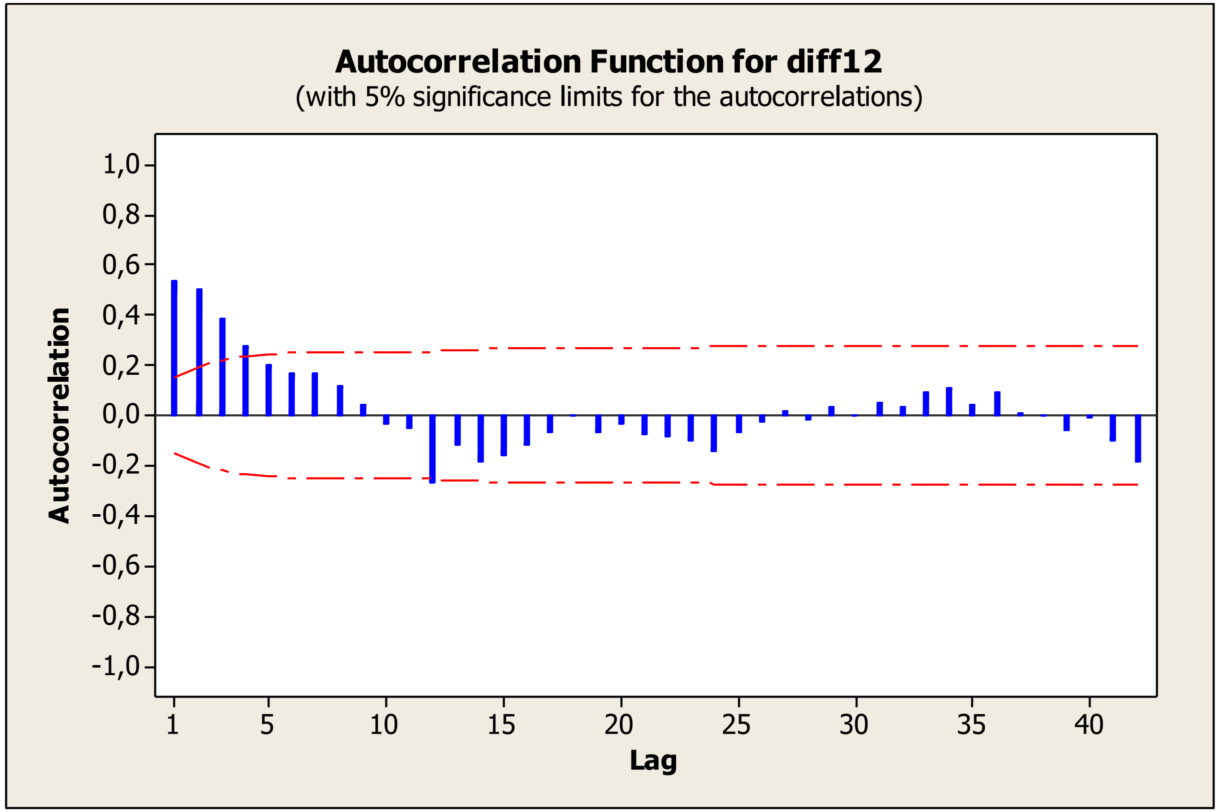
1. Säsongsrensa tidsserien med klassisk komponentuppdelning
2. Differentiera för säsong

2. Diffa för säsong. Diff12



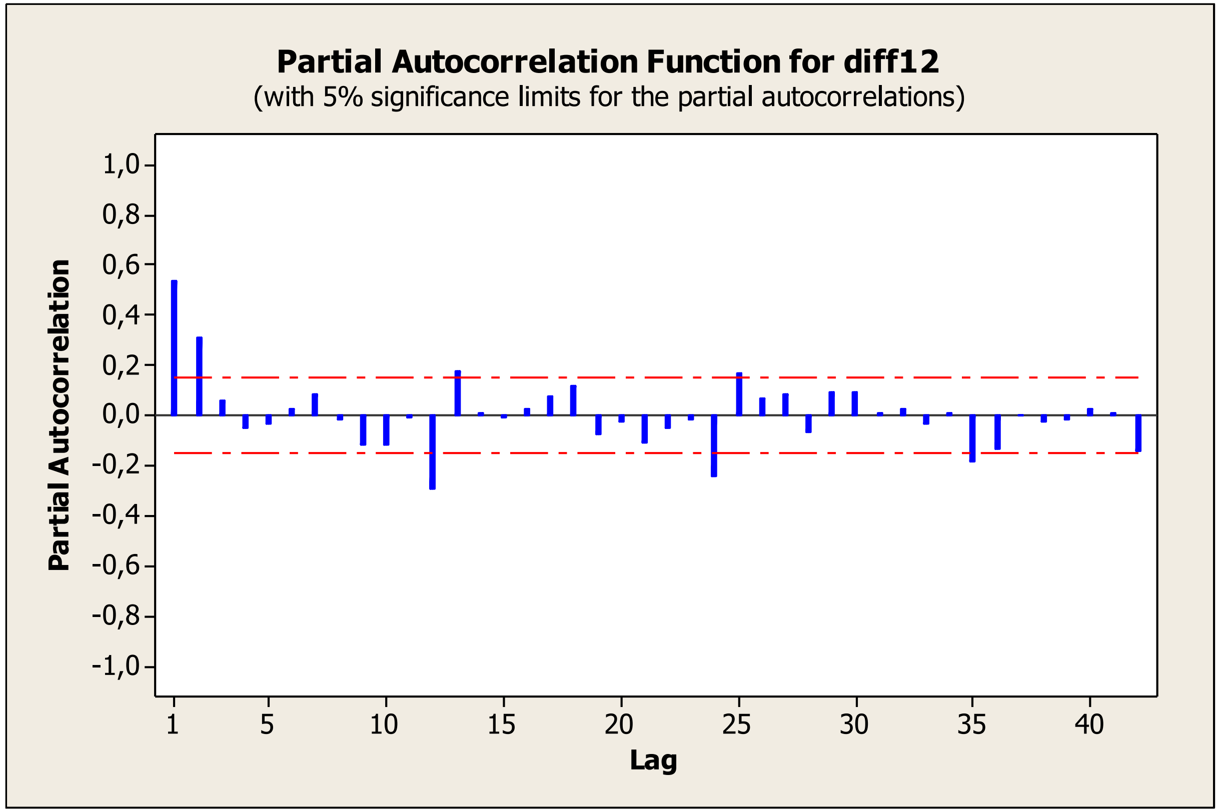
Vi ser att när vi differentierar för säsong så försvinner även trenden. Så vid **kraftig** säsongsvariation så börja alltid med att diffa för säsong.

SAC på diff12



Graf och SAC visar på att diff12 är stationär. Fortsätt då att studera SPAC.
Vi ser i SAC en spik i lag 12. Och snabbt avtagande i de första laggarna.

SPAC på diff12



Måndasdata. Vi ser spikar vid lag 12, 24, 36,…
Två spikar i början

Hur hittar vi en SARIMA-modell

När vi ska hitta en modell för säsongsdelen så gör vi på samma sätt som när vi hittade en ARMA(p,d,q)-modell.

Men nu tittar vi endast på var L:te lag.

Ex 12, 24, 36, 48,… för månadsdata och 4, 8, 12, 16,… för kvartalsdata

ARIMA(p,d,q)$(P,D,Q)_L$ Denna typ av modell kallar vi SARIMA

Förslag på modell för diff12

SAC avtar ganska snabbt i början

SPAC har två spikar i början.

→ ARMA(2,0) för diff12

→ ARIMA(2,0,0)(0,1,0) för y_t

Pröva nu denna modell så ska vi se att mönstret var 12:e spik kommer finnas kvar i SAC för residualerna.

Alternativt är att man modellerar säsongeffekterna direkt.

ARIMA Model: diff12

Estimates at each iteration

Iteration	SSE	Parameters		
0	3827,36	0,100	0,100	2,350
1	3171,55	0,250	0,213	1,562
2	3002,96	0,363	0,304	0,946
3	3002,14	0,368	0,312	0,887
4	3002,14	0,368	0,313	0,881
5	3002,14	0,368	0,313	0,881

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

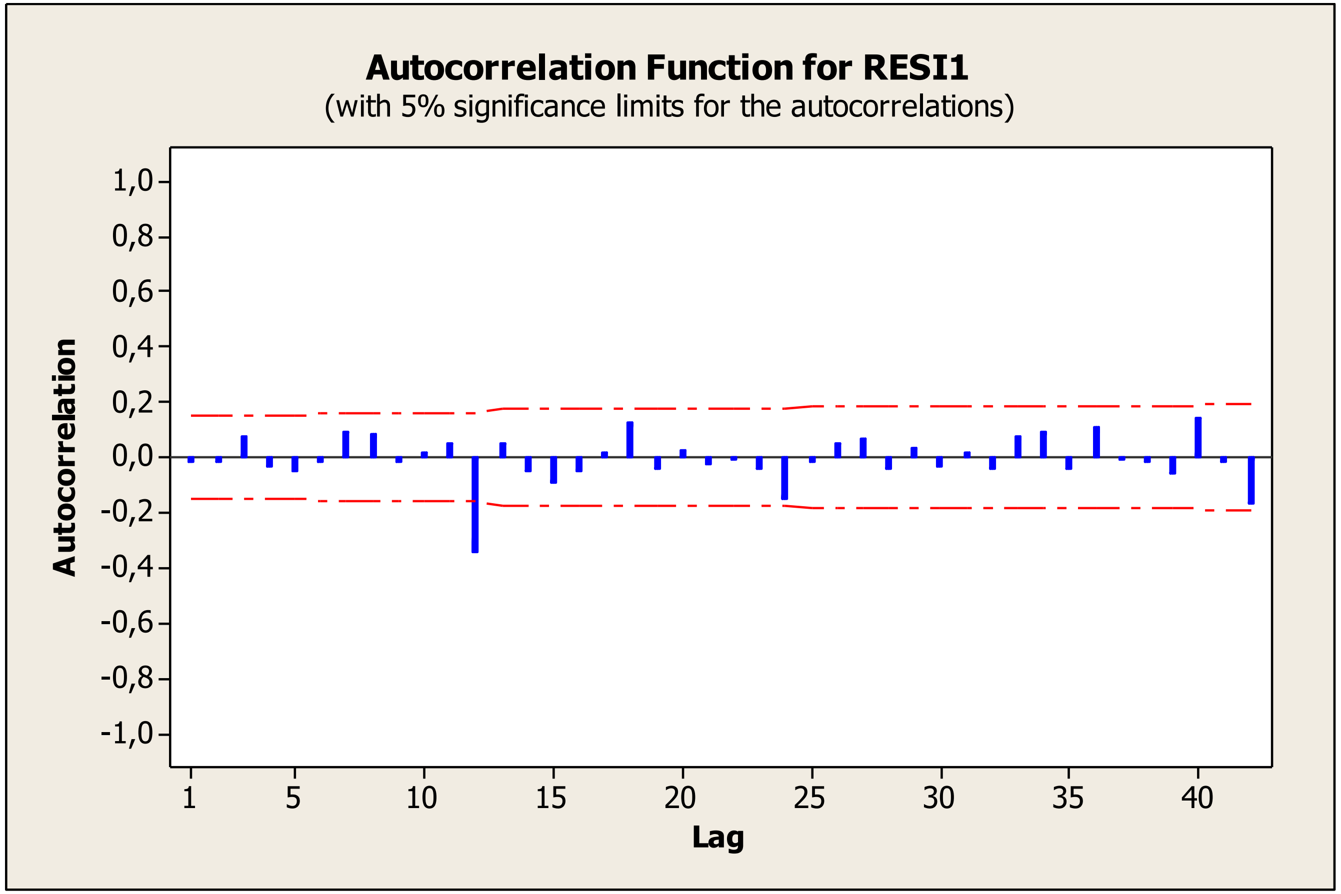
Type	Coef	SE Coef	T	P
AR 1	0,3676	0,0747	4,92	0,000
AR 2	0,3133	0,0748	4,19	0,000
Constant	0,8807	0,3328	2,65	0,009
Mean	2,760	1,043		

```
Number of observations:  166
Residuals:    SS =  2996,99 (backforecasts excluded)
              MS =  18,39  DF = 163


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              12      24      36      48
Chi-Square     26,2    37,5    45,9    64,6
DF                9      21      33      45
P-Value       0,002   0,015   0,067   0,029
```

SAC på residualerna



Fortsätt nu att modellera säsongsvariationen

- Men först ska vi titta på ett utvärderingsmått för residualerna
- Vi tar även lite teori

Box-Pierce-Statistika

Vid residualanalys så vill vi pröva om de K första autokorrelationerna för residualerna är 0.

H_0: De K första autokorrelationerna är 0

Detta görs med statistikan:

$$Q = (n-d)\sum_{l=1}^{K} r_l^2$$

Förkasta nollhypotesen om $Q > \chi^2(K-q)$

där q är antalet parametrar i modellen.

Eftersom vi vill att SAC för residualerna ska vara små så vill vi inte förkasta nollhypotesen.

Ex Pappersproduktion Första modellen

ARIMA Model: DESE1

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	0,6522	0,0577	11,31	0,000
Constant	0,21802	0,09463	2,30	0,022

Differencing: 1 regular difference

Number of observations: Original series 178, after differencing 177

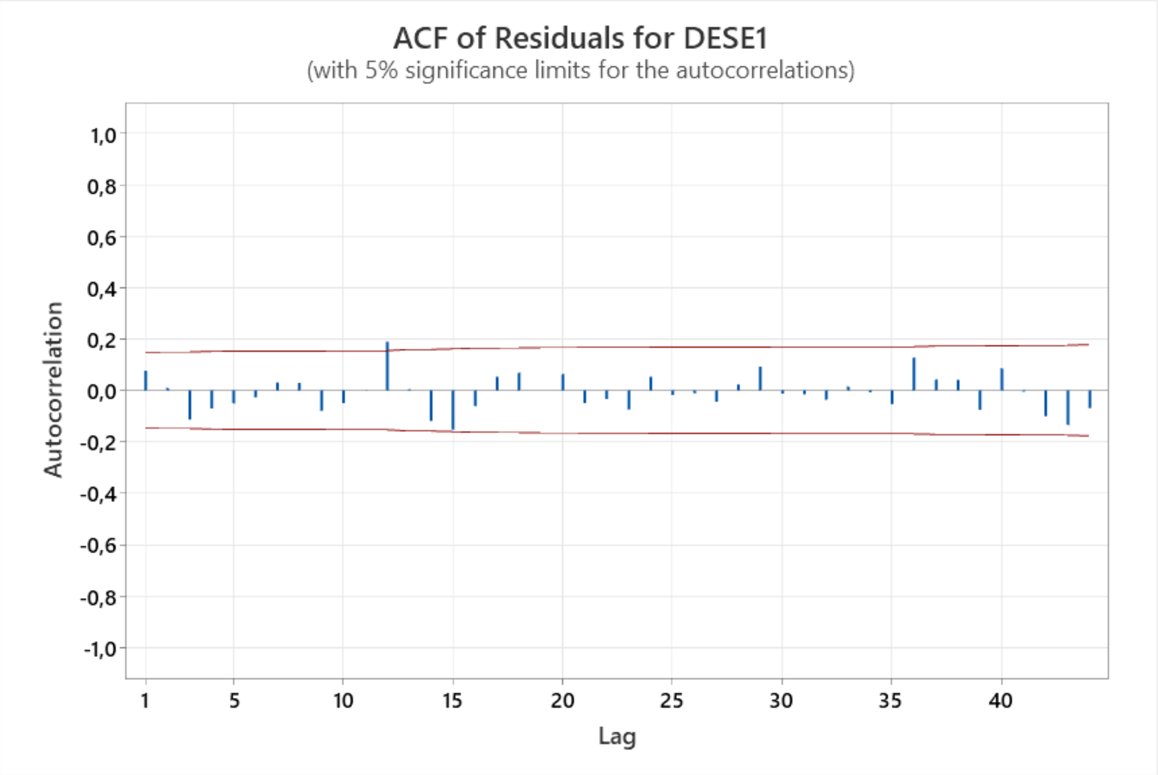
Residuals: SS = 2273,62 (backforecasts excluded)

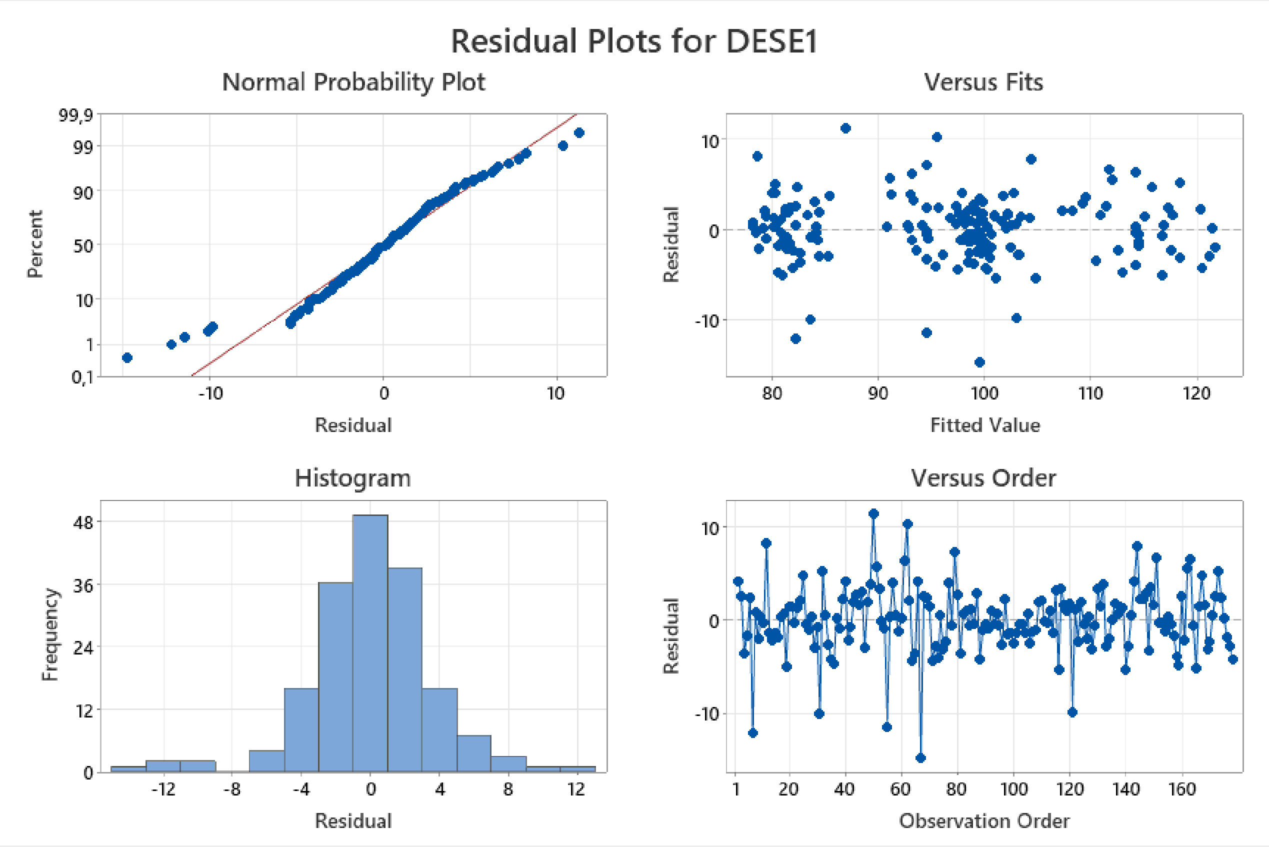
MS = 12,99 DF = 175

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	13,8	26,7	33,8	48,9
DF	10	22	34	46
P-Value	0,182	0,223	0,476	0,357

Residualanalys Första modellen





Andra modellen

ARIMA Model: diff12

Estimates at each iteration

Iteration	SSE	Parameters		
0	3827,36	0,100	0,100	2,350
1	3171,55	0,250	0,213	1,562
2	3002,96	0,363	0,304	0,946
3	3002,14	0,368	0,312	0,887
4	3002,14	0,368	0,313	0,881
5	3002,14	0,368	0,313	0,881

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,3676	0,0747	4,92	0,000
AR 2	0,3133	0,0748	4,19	0,000
Constant	0,8807	0,3328	2,65	0,009
Mean	2,760	1,043		

```
Number of observations:  166
Residuals:     SS =  2996,99 (backforecasts excluded)
               MS =  18,39  DF = 163


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              12      24      36      48
Chi-Square     26,2    37,5    45,9    64,6
DF                9      21      33      45
P-Value       0,002   0,015   0,067   0,029
```

Inte bra. p-värdena är små

Teori SARIMA= SäsongsARIMA ARIMA(p,d,q)(P,D,Q)L

I en ren SARIMA modellerar vi endast var $L\!:\!te$ tidsavstånd. Låt z_t vara den stationära serien, dvs den eventuellt differentierade serien.

SAR(P) L är säsongslängd: 2, 4, 12 som exempel

$$z_t = \delta + \phi_{1,L} z_{t-L} + \phi_{2,L} z_{t-2L} + \dots + \phi_{P,L} z_{t-PL} + a_t$$

$$z_t - \phi_{1,L} z_{t-L} - \phi_{2,L} z_{t-2L} - \dots - \phi_{P,L} z_{t-PL} = \delta + a_t$$

$$(1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \dots - \phi_{P,L} B^{PL}) z_t = \delta + a_t$$

$$\Phi_P^*(B^L) z_t = \delta + a_t$$

$$\Phi_P^*(B^L) = (1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \dots - \phi_{P,L} B^{PL})$$

P väljs på samma sätt som tidigare men nu studeras endast var $L\!:\!te$ spik i SAC och SPAC

SMA(Q), L är säsongslängd: 2, 4, 12

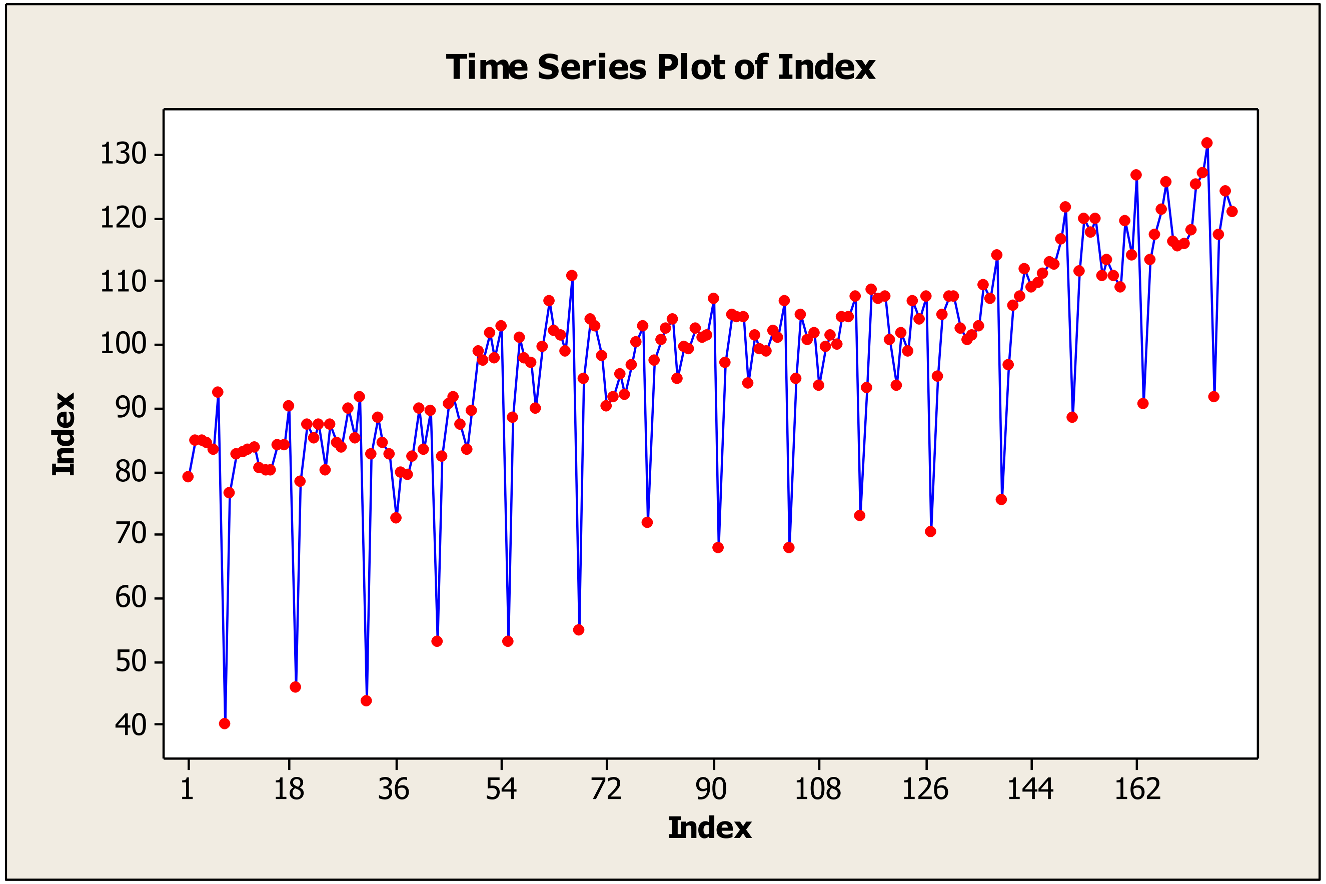
$$z_t = \delta + a_t - \theta_{1,L} a_{t-L} - \theta_{2,L} a_{t-2L} - \ldots - \theta_{Q,L} a_{t-QL}$$

$$z_t = \delta + (1 - \theta_{1,L} B^L - \theta_{2,L} B^{2L} - \ldots - \theta_{Q,L} B^{QL}) a_t$$

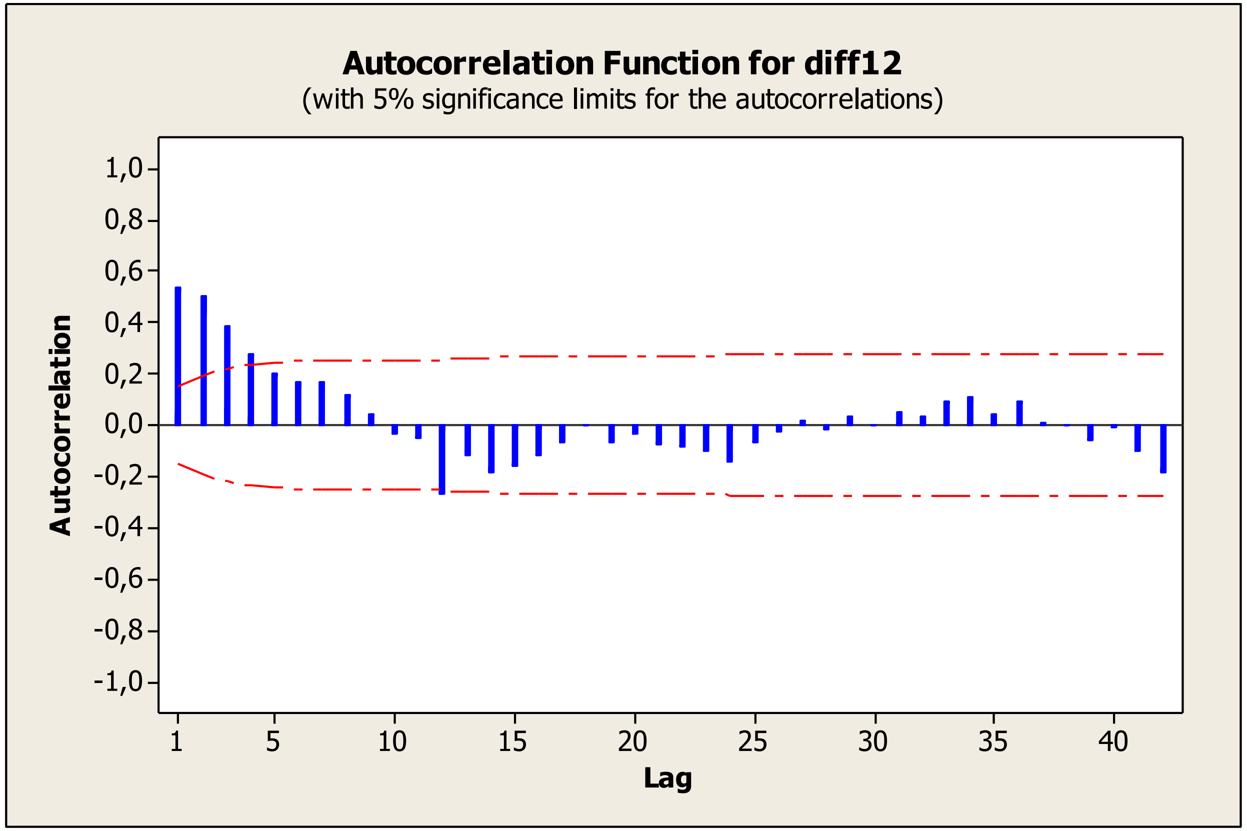
$$z_t = \delta + \Theta_Q^*(B^L) a_t$$

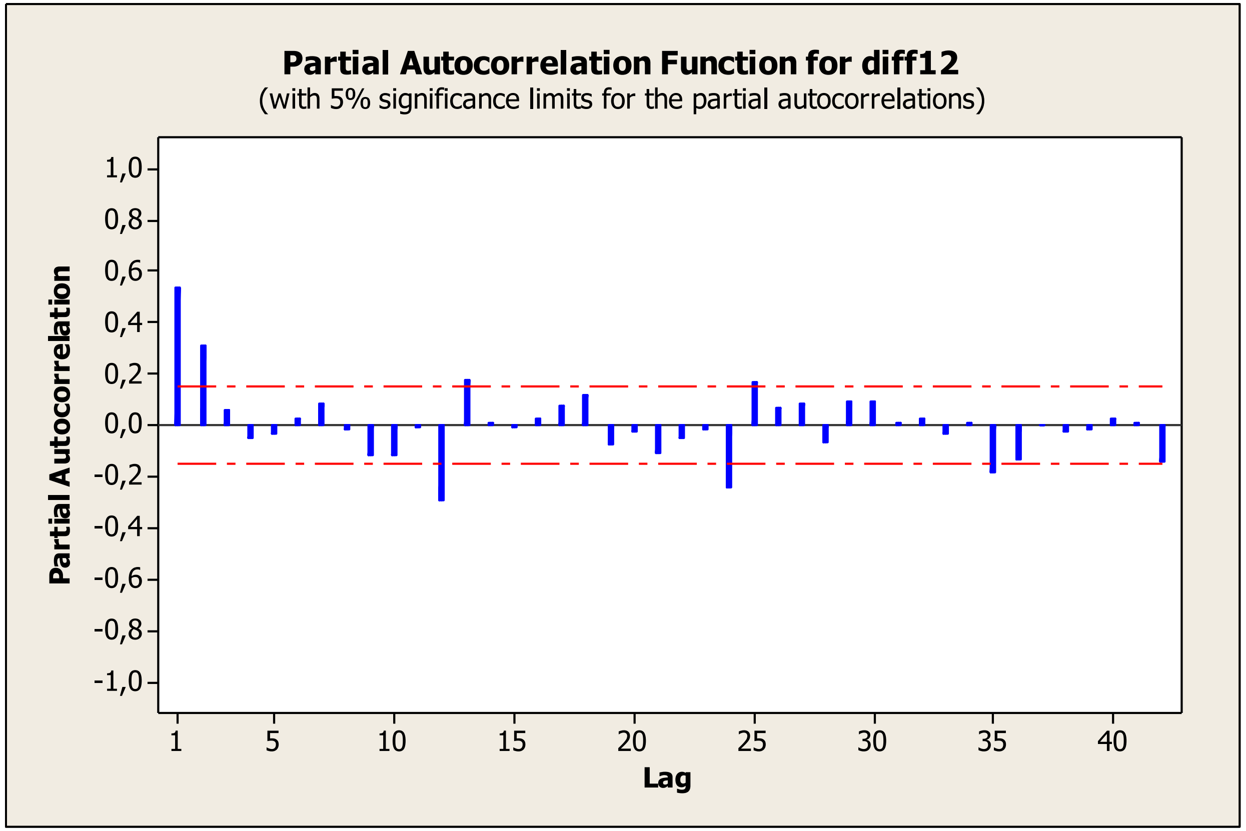
Q väljs på samma sätt som tidigare men vi tittar endast på var $L{:}te$ spik i SAC och SPAC

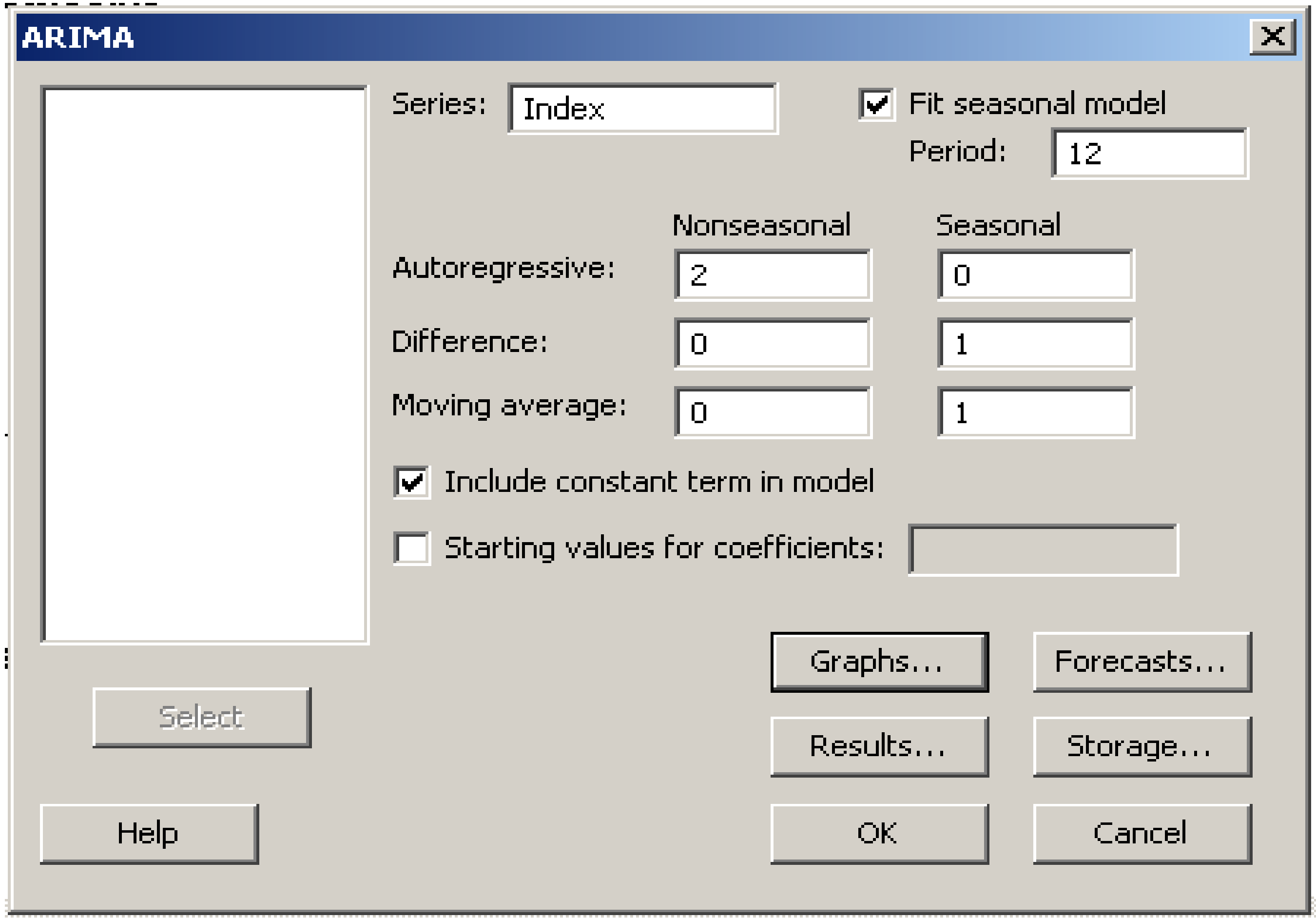
Ex Pappersproduktion



Differentiera serien för säsong, diff12, D=1







Vi prövar med att anpassa en ARIMA(2,0,0)(0,1,1)$_{12}$

ARIMA Model: Index

Estimates at each iteration

Iteration	SSE	Parameters			
0	3629,29	0,100	0,100	0,100	2,350
1	2779,04	0,250	0,213	0,242	1,542
2	2446,76	0,337	0,272	0,392	1,102
3	2273,36	0,394	0,300	0,542	0,844
4	2190,27	0,433	0,300	0,678	0,726
5	2178,22	0,444	0,282	0,721	0,744
6	2176,33	0,447	0,275	0,737	0,756
7	2175,96	0,447	0,272	0,744	0,761
8	2175,87	0,448	0,271	0,747	0,763
9	2175,86	0,448	0,271	0,749	0,764
10	2175,85	0,448	0,271	0,750	0,764
11	2175,85	0,448	0,270	0,750	0,765

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0,4480	0,0752	5,96	0,000
AR	2	0,2704	0,0754	3,59	0,000
SMA	12	0,7500	0,0564	13,30	0,000
Constant		0,76464	0,07526	10,16	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 178, after differencing 166

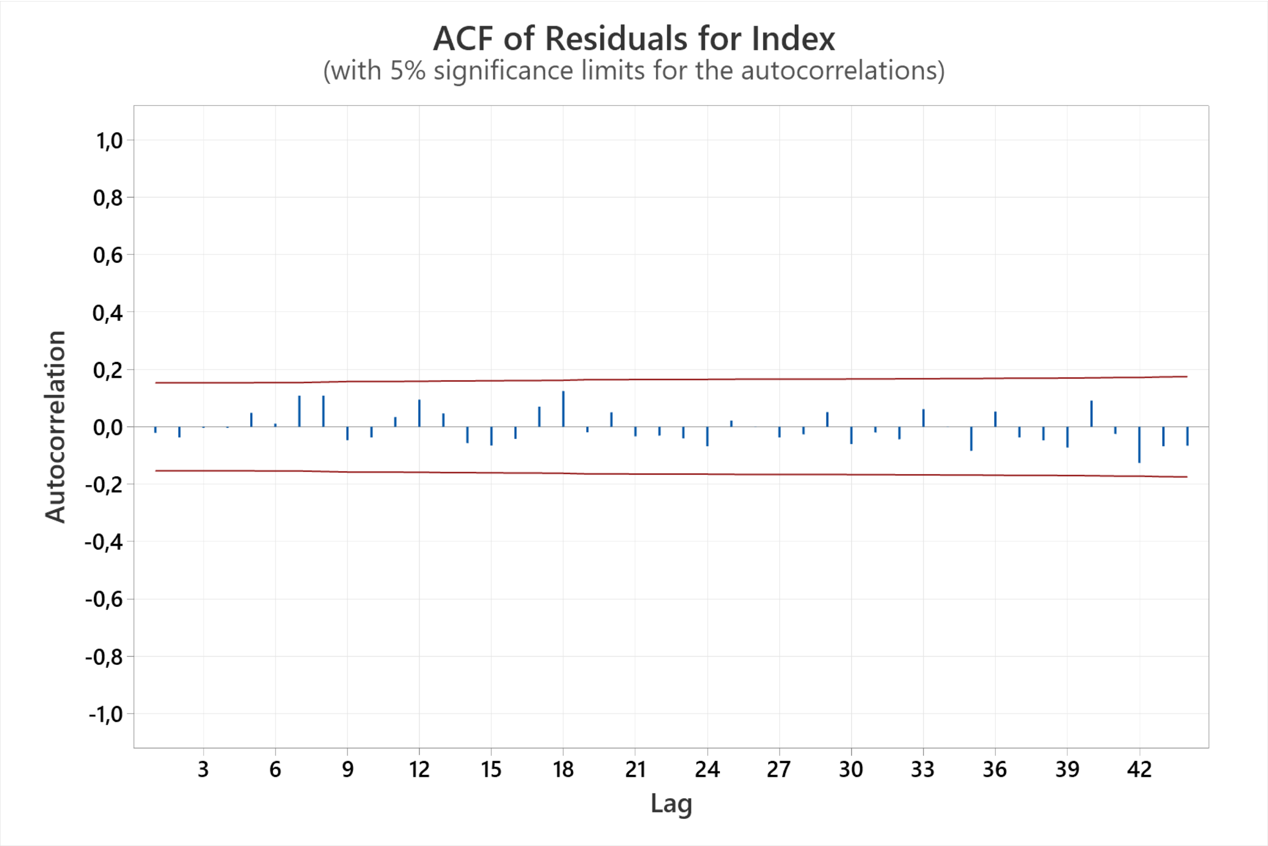
Residuals: SS = 2080,86 (backforecasts excluded)

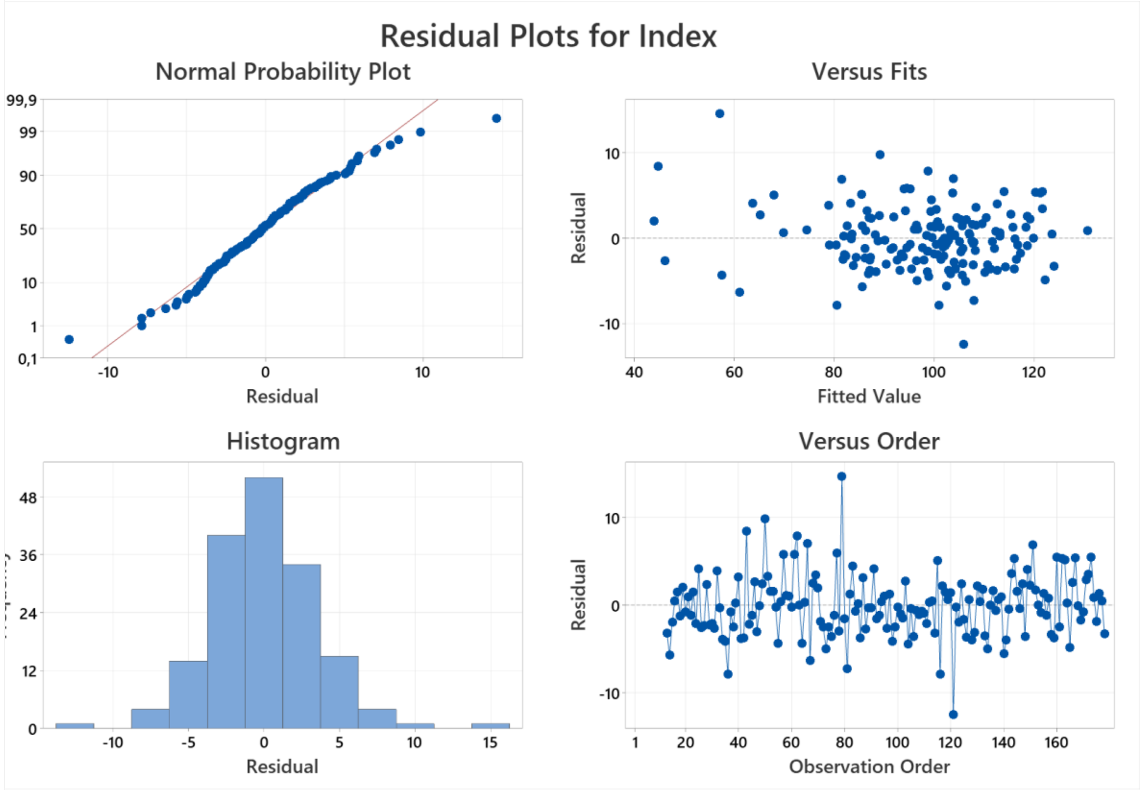
MS = 12,84 DF = 162

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	6,3	14,4	19,9	35,1
DF	8	20	32	44
P-Value	0,614	0,809	0,954	0,828

Residualanalys Andra modellen





Prognos

Gör först prognos för den differentierade serien eller för den säsongsrensade serien.

Därefter justeras prognosen så att den gäller för y_t

Notation:

$\hat{Y}_{T+\tau}$*(T)* = prognos för $Y_{T+\tau}$ då prognosen görs vid tidpunkt T för tidpunkt $T + \tau$

Prognos för z_t och a_t då de är inne i modellen

- Om t är dåtid så prognostiseras z_t med observation och a_t med residual
- Om t är framtid så prognostiseras z_t med prognos och a_t med 0

1. Säsongsrensad serie, MA(1)

Beräkna prognoser för y_t för nästkommande tre månader.

Visas på tavlan

De sista 5 värdena i serien

DESE	resid	År månad	
122,741	2,32237	2004	6
121,552	0,10709	2004	7
119,760	-1,93984	2004	8
118,325	-2,91865	2004	9
116,160	-4,2862	2004	10

Seasonal Indices

Period	Index
1	0,3795
2	0,8295
3	0,7441
4	5,3920
5	3,6420
6	8,9587
7	-29,9517
8	-2,5601
9	5,8753
10	4,6399
11	5,1753
12	-3,1247

Prognos för nov dec 2004 samt jan 2005 enligt

Forecasts from period 178

Period	Forecast	95% Limits Lower	Upper	Actual
179	119,174	112,107	126,240	
180	119,392	111,910	126,873	
181	119,610	111,735	127,484	

Ex Pappersproduktion Första modellen

ARIMA Model: DESE1

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	0,6522	0,0577	11,31	0,000
Constant		0,21802	0,09463	2,30	0,022

Differencing: 1 regular difference

Number of observations: Original series 178, after differencing 177

Residuals: SS = 2273,62 (backforecasts excluded)

MS = 12,99 DF = 175

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
=Chi-Square	13,8	26,7	33,8	48,9
DF	10	22	34	46
P-Value	0,182	0,223	0,476	0,357

$$(1-B)DESE_t = \delta + \Theta_1(B)a_t$$

2. Modell 2

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,4480	0,0752	5,96	0,000
AR 2	0,2704	0,0754	3,59	0,000
SMA 12	0,7500	0,0564	13,30	0,000
Constant	0,7646	0,0753	10,16	0,000

```
Differencing: 0 regular, 1 seasonal of order 12
Number of observations:  Original series 178, after
differencing 166
Residuals:    SS =  2080,86 (backforecasts excluded)
              MS =  12,84  DF = 162
```

$$\Phi_2(B)(1-B^{12})y_t = \delta + \Theta_1^*(B^{12})a_t$$

De sista 13 värdena i serien

Beräkna prognos för y_t för nästkommande tre månader. Visas på tavlan. Jämför prognoserna med modellen ovan

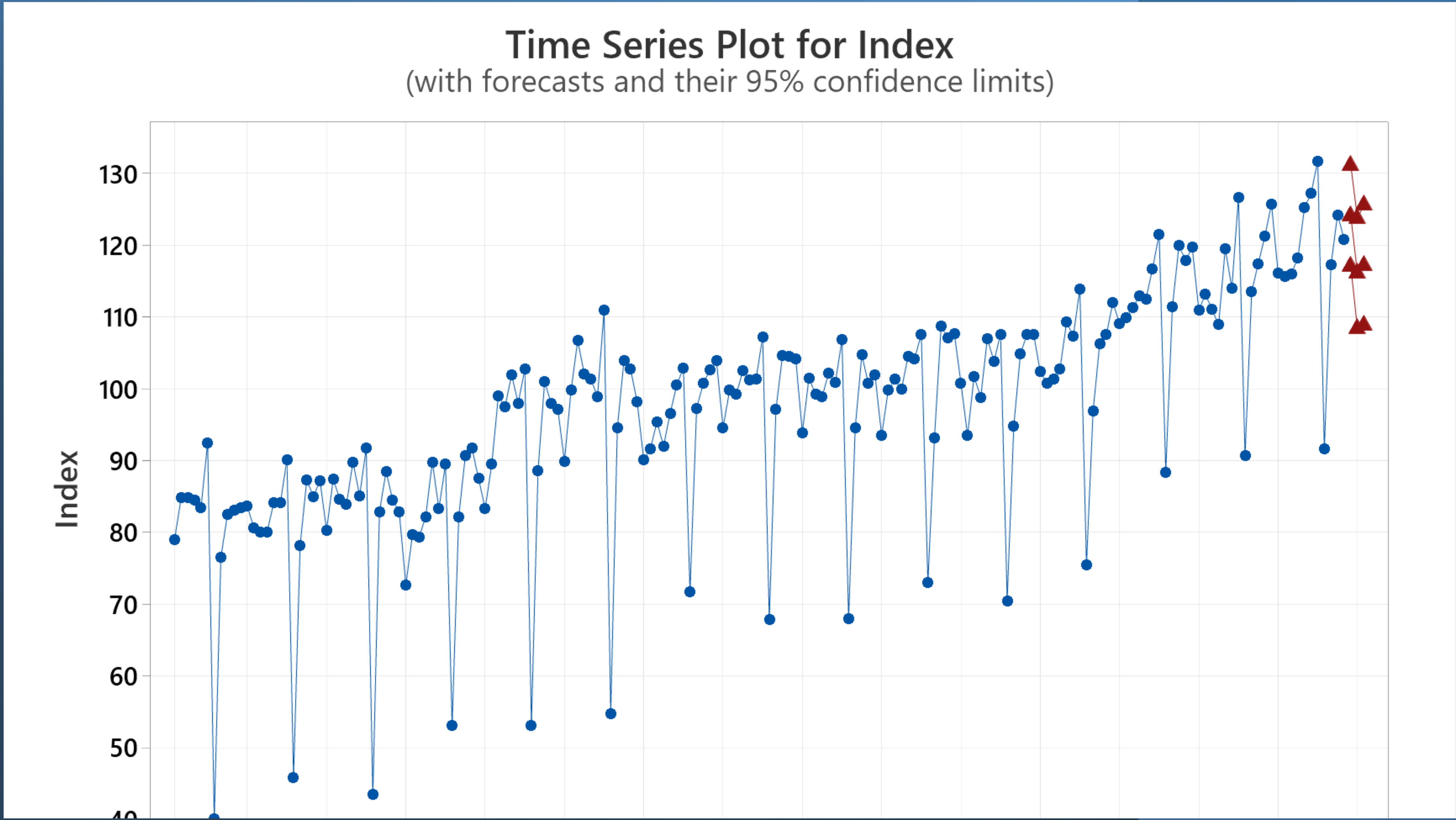
År	mån	y_t	diff12	Resid
2003	10	121,2	3,4	2,55181
2003	11	125,7	6,0	5,38271
2003	12	116,1	5,2	-0,06830
2004	1	115,6	2,4	-1,71964
2004	2	116,0	5,0	-0,75025
2004	3	118,2	9,3	2,85530
2004	4	125,2	5,7	3,49452
2004	5	127,2	13,2	5,49363
2004	6	131,7	5,1	0,87310
2004	7	91,6	0,9	-1,85568
2004	8	117,2	3,7	1,32046
2004	9	124,2	6,8	0,50082
2004	10	120,8	-0,4	-3,29763

ARIMA(2,0,0)(0,1,1)$_{12}$

- Prognoser enligt MINITAB

Forecasts from period 178

		95% Limits		
Period	**Forecast**	**Lower**	**Upper**	**Actual**
179	124,087	117,062	131,113	
180	116,085	108,387	123,784	
181	117,212	108,832	125,592	


Time Series Plot for Index
(with forecasts and their 95% confidence limits)
Index
130
120
110
100
90
80
70
60
50

Ett annat Exempel $\text{SARIMA}(p,d,q)(P,D,Q)_L = \text{SARIMA}(1,1,0)(0,1,1)_{12}$

$\text{SARIMA}(0,1,0)(0,1,0)_{12}$

$(1-B)(1-B^{12})y_t$ d=1 och D=1

$\text{SARIMA}(1,0,0)(0,0,0)_{12}$

$(1-\varphi_1 B)y_t = a_t$ p=1

$\text{SARIMA}(0,0,0)(0,0,1)_{12}$

$y_t = (1-\theta_{1,12}B^{12})a_t$ Q=1

Sätt ihop dessa tre ekvationer till en $\text{SARIMA}(1,1,0)(0,1,1)_{12}$

$$(1-\varphi_1 B)(1-B)(1-B^{12})y_t = (1-\theta_{1,12}B^{12})a_t$$

$$\Phi_1(B)(1-B)(1-B^{12})y_t = \Theta_1^*(B^{12})a_t$$

Största modellen ARIMA(p,d,q)(P,D,Q)L

$$\Phi_p(B)\Phi_P^*(B^L)(1-B)^d(1-B^L)^D y_t = \Theta_q(B)\Theta_Q^*(B^{12})a_t$$