



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2010-06-04
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	KÅRA
Tid	8-12
Kurskod	732G25
Provkod	TENT
Kursnamn/benämning Provnamn/benämning	Statistisk analys av samhällsdata Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	9,50
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 28 14 63, carita.lilja@liu.se
Tillåtna hjälpmedel	Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: <i>Forecasting, Time Series and Regression</i> , alla upplagor tillåtna, som inte får innehålla anteckningar men får ha markeringar och flärpar.
Övrigt	G=12, VG=16
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	30

Tentamen i Statistisk analys av samhällsdata, 2010-06-04

Skrivtid: kl: 8-12

Hjälpmedel: Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: *Forecasting, Time Series and Regression*, alla upplagor tillåtna, som inte får innehålla anteckningar men får ha markeringar och flärpar.

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

Data i denna tenta är hämtade från SCB

1

Vi ska nu studera befolkningsstatistik för Östergötland år 2006.

region	ålder	kön	tid	Folkmängden den 1 november
05 Östergötlands län	0-4 år	kvinnor	2006	10639
05 Östergötlands län	5-9 år	kvinnor	2006	10182
05 Östergötlands län	10-14 år	kvinnor	2006	13122
05 Östergötlands län	15-19 år	kvinnor	2006	14012
05 Östergötlands län	20-24 år	kvinnor	2006	12864
05 Östergötlands län	25-29 år	kvinnor	2006	12025
05 Östergötlands län	30-34 år	kvinnor	2006	12740
05 Östergötlands län	35-39 år	kvinnor	2006	13390
05 Östergötlands län	40-44 år	kvinnor	2006	14188
05 Östergötlands län	45-49 år	kvinnor	2006	12824
05 Östergötlands län	50-54 år	kvinnor	2006	12888
05 Östergötlands län	55-59 år	kvinnor	2006	14023
05 Östergötlands län	60-64 år	kvinnor	2006	13756
05 Östergötlands län	65-69 år	kvinnor	2006	10276
05 Östergötlands län	70-74 år	kvinnor	2006	8721
05 Östergötlands län	75-79 år	kvinnor	2006	8249
05 Östergötlands län	80-84 år	kvinnor	2006	7428
05 Östergötlands län	85-89 år	kvinnor	2006	5157
05 Östergötlands län	90-94 år	kvinnor	2006	2064
05 Östergötlands län	95-99 år	kvinnor	2006	488
05 Östergötlands län	100+ år	kvinnor	2006	59

region	tid	Folkmängden den 1 november
05 Östergötlands län	2006	417836

region	tid	Levande födda
05 Östergötlands län	2006	4575

region	moderns ålder	tid	Levande födda
05 Östergötlands län	-14 år	2006	0
05 Östergötlands län	15 år	2006	1
05 Östergötlands län	16 år	2006	5
05 Östergötlands län	17 år	2006	8
05 Östergötlands län	18 år	2006	17
05 Östergötlands län	19 år	2006	31
05 Östergötlands län	20 år	2006	65
05 Östergötlands län	21 år	2006	73
05 Östergötlands län	22 år	2006	105

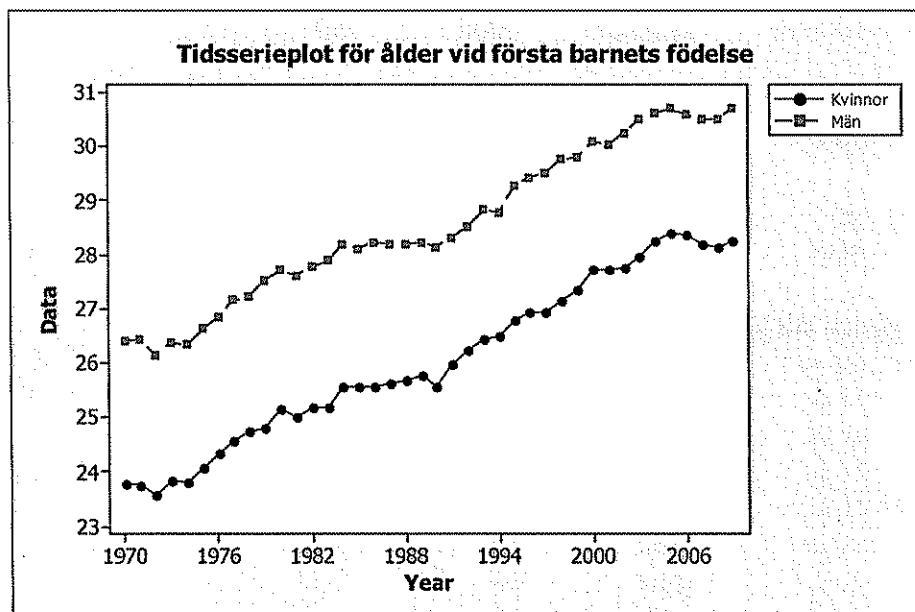
05 Östergötlands län	23 år	2006	140
05 Östergötlands län	24 år	2006	183
05 Östergötlands län	25 år	2006	190
05 Östergötlands län	26 år	2006	214
05 Östergötlands län	27 år	2006	271
05 Östergötlands län	28 år	2006	308
05 Östergötlands län	29 år	2006	363
05 Östergötlands län	30 år	2006	358
05 Östergötlands län	31 år	2006	364
05 Östergötlands län	32 år	2006	355
05 Östergötlands län	33 år	2006	315
05 Östergötlands län	34 år	2006	302
05 Östergötlands län	35 år	2006	238
05 Östergötlands län	36 år	2006	178
05 Östergötlands län	37 år	2006	139
05 Östergötlands län	38 år	2006	115
05 Östergötlands län	39 år	2006	80
05 Östergötlands län	40 år	2006	66
05 Östergötlands län	41 år	2006	46
05 Östergötlands län	42 år	2006	17
05 Östergötlands län	43 år	2006	17
05 Östergötlands län	44 år	2006	7
05 Östergötlands län	45 år	2006	1
05 Östergötlands län	46 år	2006	1
05 Östergötlands län	47 år	2006	2
05 Östergötlands län	48 år	2006	0
05 Östergötlands län	49+ år	2006	0

Lös uppgifterna nedan med hjälp av uppgifterna ovan

- Beräkna allmänna födelsetalet, crude birth rate 1p
- Beräkna generella födelsetalet, general birth rate 1p
- Beräkna den summerade fruktsamhetskvoten, TFR. Tolka 2p

2

I grafen nedan visas medelåldern vid första barnets födelse för män respektive kvinnor åren 1970 till 2009



För att skatta hur stor medelåldern ökar per år så har en regressionsmodell anpassats. Med hjälp av modellen ville man också skatta ålderskillnaden mellan kvinnor och män. Därför har kön inkluderats i modellen som en dummyvariabel där kvinnor har fått koden 0 och män koden 1.

Resultat:

Regression Analysis: Medelåldern vid första barnets versus år; dummy

The regression equation is

$$\text{Medelåldern vid första barnets} = 23,5 + 0,126 \text{ år} + 2,50 \text{ dummy}$$

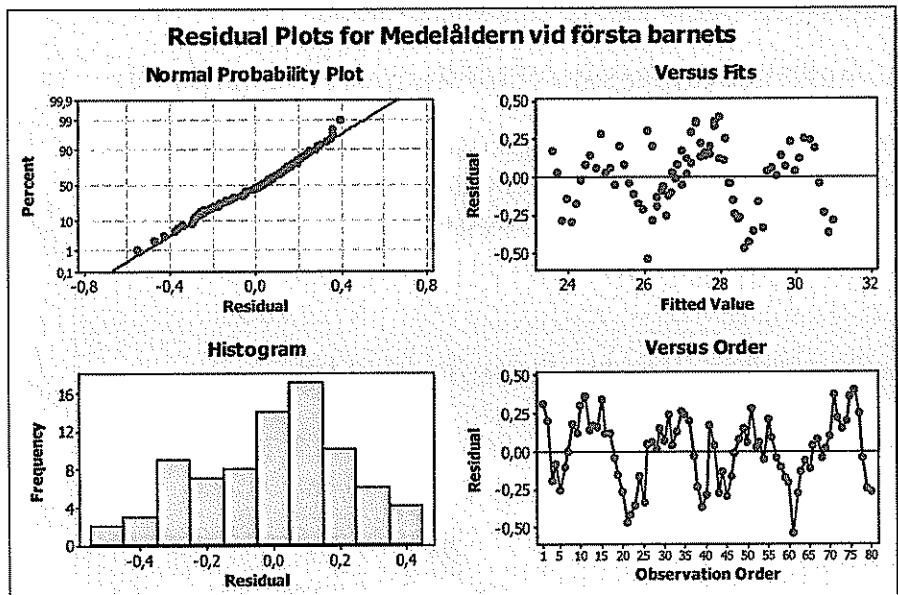
Predictor	Coef	SE Coef
Constant	23,4670	0,0557
år	0,125745	0,002126
dummy	2,50275	0,04908

S = 0,219470

Analysis of Variance

Source	DF	SS	MS
Regression	2	293,83	146,91
Residual Error	77	3,71	0,05
Total	79	297,54	

Durbin-Watson statistic = 0,596979



- Hur många månader har medelåldern vid första barnets födelse ökat i snitt per år enligt modellen? Hur stor är ålderskillnaden mellan mäns och kvinnors medelålder vid första barnets födelse i snitt enligt modellen? 2p
- Prediktera medelåldern för kvinnor år 2010 med hjälp av modellen ovan. 1p
- Pröva på 5% signifikansnivå om regressionskoefficienterna för år och dummy är signifikant skilda från 0. Gör ett test för vardera parameter. 1p
- Beräkna förklaringsgraden R^2 och tolka den. 1p
- Tolka resultatet av Durbin-Watson statistikan. 1p
- Förklara vad en residual är och tolka de 4 graferna för residualerna ovan. 1p

3

Utjämnna följande tidsserie med ett centrerat 5 punkters glidande medelvärde

4 5 3 2 5 3 4 7 10 6

2p

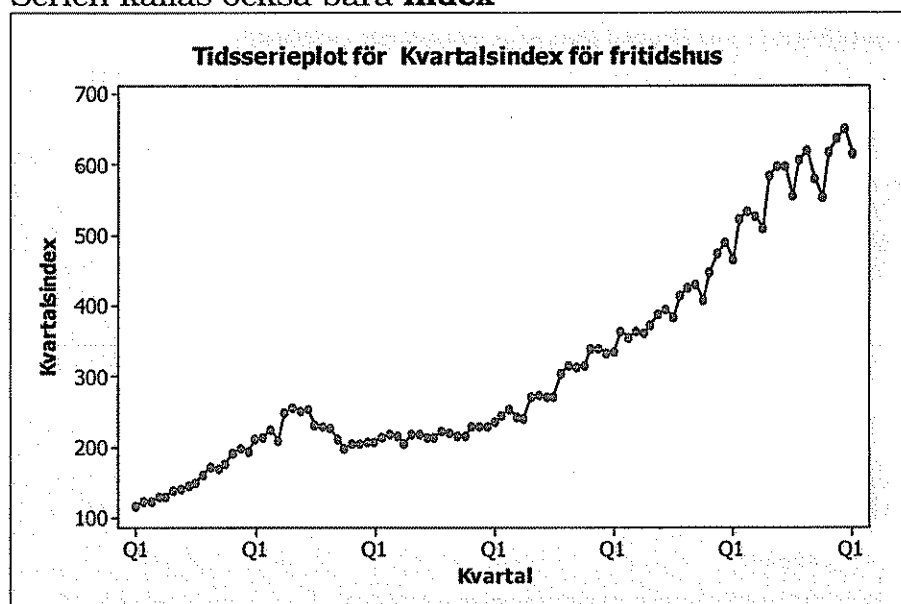
4 a-e

I grafen nedan visas

Fastighetsprisindex för fritidshus. Kvartal 1986K1-2010K1

(Q1 som syns i vissa grafer nedan betyder kvartal 1)

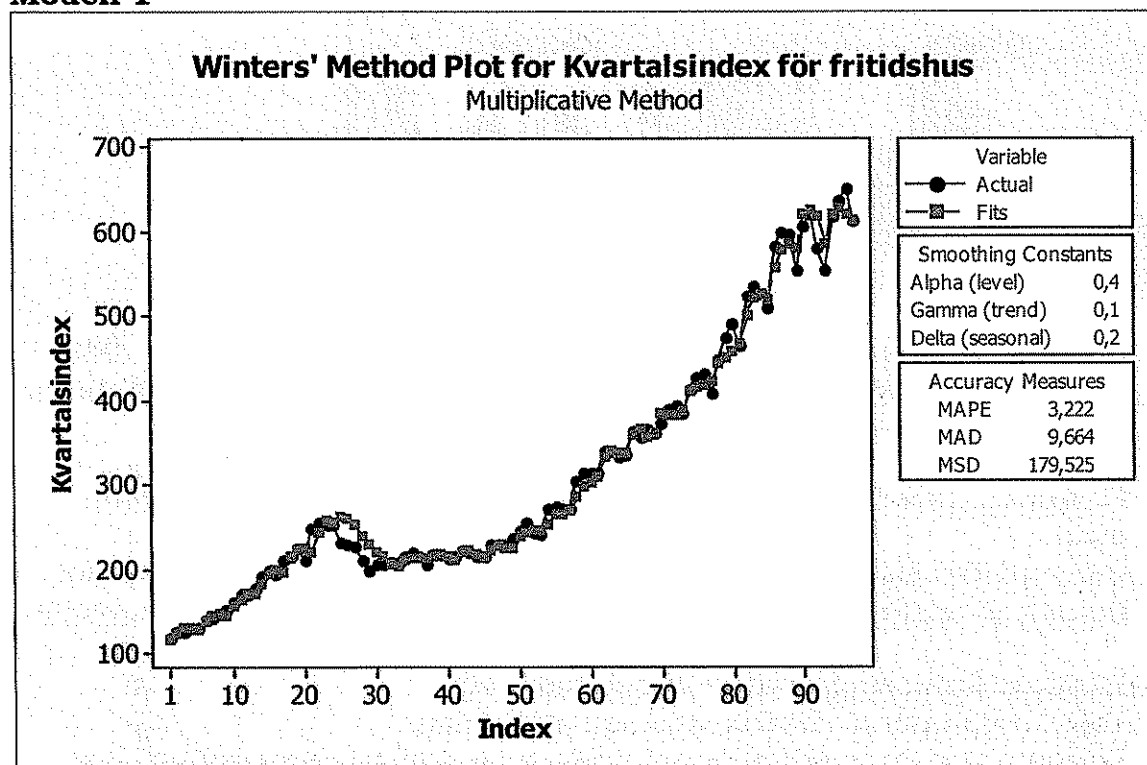
Serien kallas också bara **index**



Nedan har två typer av modeller anpassats och från modell 1 och 3 har prognoser beräknats.

- a) Studera modell 1. Visa hur modellen ser ut. Förklara hur parametrarna skattas samt hur man går tillväga för att hitta bäst anpassning. Är du nöjd med modellen. Motivera. 2p

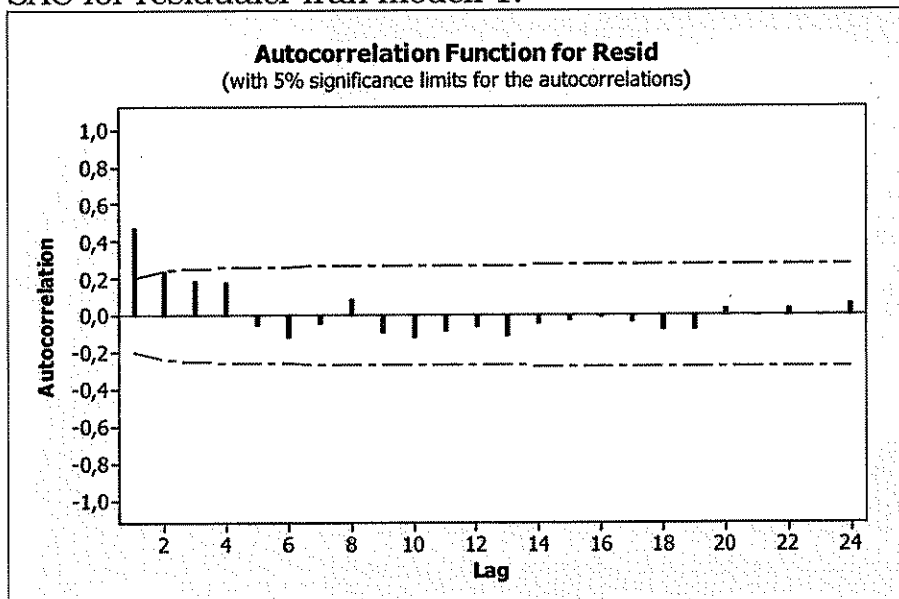
Modell 1



Forecasts

Period	Forecast	Lower	Upper
98	671,366	647,691	695,041
99	682,891	657,686	708,095
100	673,571	646,553	700,588
101	650,058	620,998	679,119

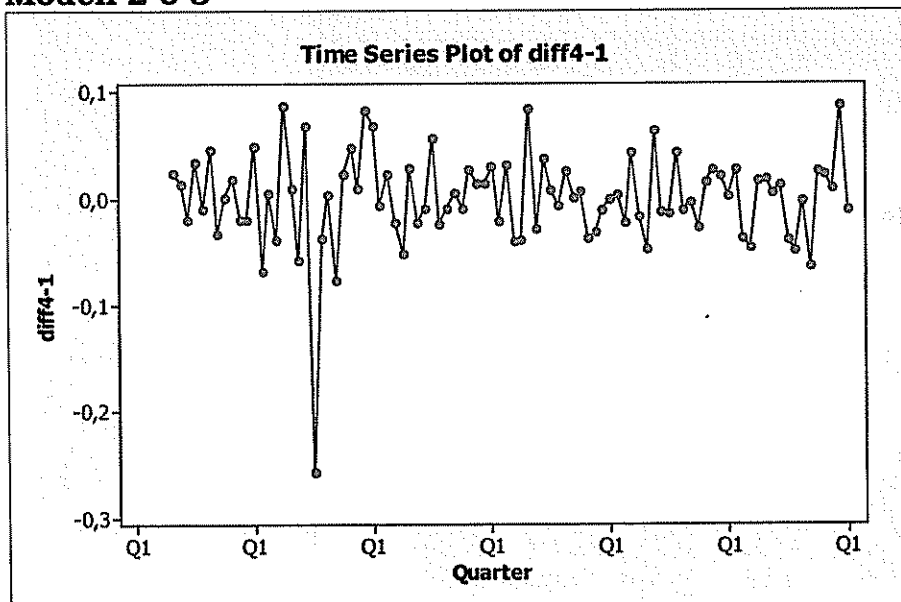
SAC för residualer från modell 1.

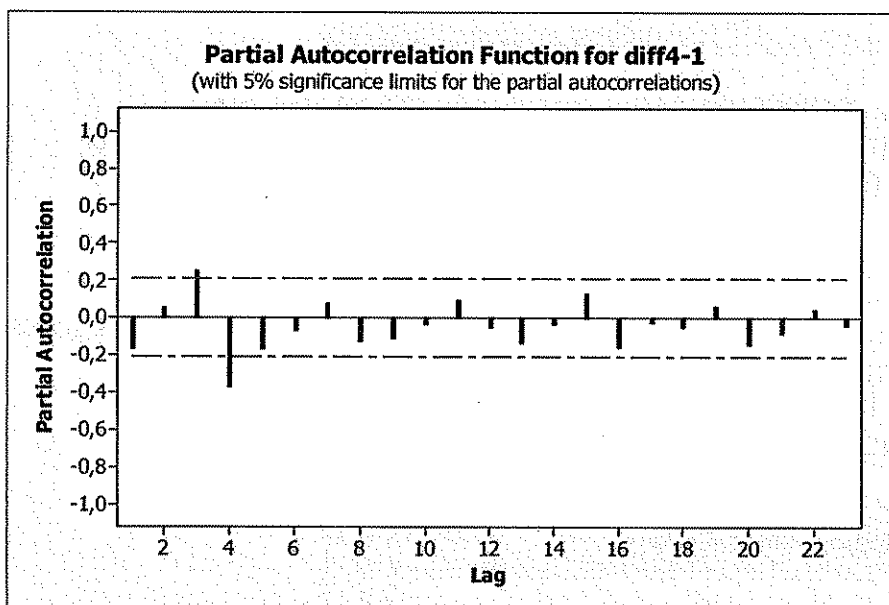
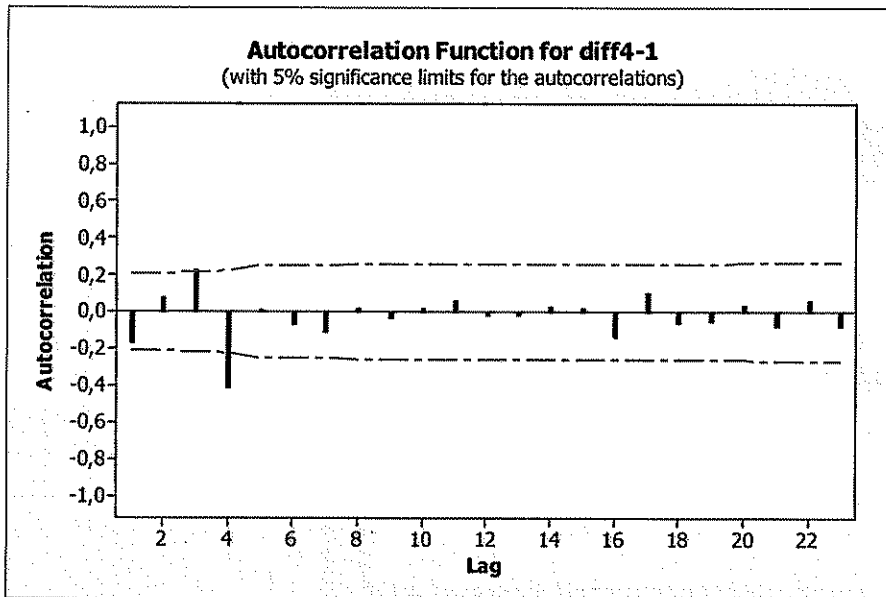


Nedan har två säsong ARIMA modeller anpassats. Först har data logaritmerats med den naturliga logaritmen. Denna serie heter **ln index**. Sedan har **ln index** differentierats för säsong och för trend. Denna serie heter **diff4-1**

- b) Förklara varför serien har logaritmerats. 1p
- c) Vilken av modell 2 o 3 anser du vara bäst. Motivera. 1p
- d) För modell 3 har prognoser för ln index beräknats. Räkna om dessa till prognoser för index. Jämför prognosintervallen från modell 1 och modell 3. Vilka föredrar du? 1p
- e) Låt $\text{diff4-1} = z_t$. Visa hur modell 3 ser ut för z_t . 2p

Modell 2 o 3





Modell 2

ARIMA Model: In index

Estimates at each iteration

Iteration	SSE		Parameters				
0	1,10237	0,100	0,100	0,100	0,100	0,081	
1	0,28629	0,088	-0,043	0,112	0,243	-0,027	
2	0,27429	-0,062	-0,034	-0,036	0,261	-0,029	
3	0,26170	-0,212	-0,024	-0,183	0,282	-0,031	
4	0,19901	-0,362	0,049	-0,316	0,413	-0,020	
5	0,17517	-0,226	0,085	-0,166	0,481	-0,013	
6	0,14776	-0,376	0,134	-0,301	0,592	-0,007	
7	0,13583	-0,243	0,175	-0,151	0,679	-0,003	
8	0,13004	-0,393	0,195	-0,295	0,750	-0,002	
9	0,12763	-0,257	0,215	-0,145	0,795	-0,001	
10	0,12626	-0,407	0,215	-0,292	0,831	-0,001	
11	0,12584	-0,269	0,228	-0,142	0,849	-0,000	
12	0,12560	-0,418	0,224	-0,292	0,864	-0,000	
13	0,12548	-0,278	0,238	-0,142	0,873	-0,000	
14	0,12543	-0,363	0,230	-0,231	0,875	-0,000	
15	0,12540	-0,311	0,235	-0,174	0,878	-0,000	
16	0,12540	-0,369	0,234	-0,235	0,880	-0,000	
17	0,12538	-0,309	0,240	-0,171	0,883	-0,000	
18	0,12537	-0,343	0,239	-0,207	0,884	-0,000	
19	0,12536	-0,339	0,243	-0,202	0,886	-0,000	
20	0,12536	-0,339	0,243	-0,202	0,886	-0,000	

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type		Coef	SE Coef
AR	1	-0,3391	0,7272
SAR	4	0,2429	0,1517
MA	1	-0,2025	0,7617
SMA	4	0,8857	0,0914
Constant		-0,0001336	0,0006898

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 97, after differencing 92

Residuals: SS = 0,122825 (backforecasts excluded)
MS = 0,001412 DF = 87

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

Lag	12	24	36	48
Chi-Square	6,0	10,2	21,9	33,0
DF	7	19	31	43
P-Value	0,541	0,947	0,885	0,865

Modell 3

ARIMA Model: In index

Estimates at each iteration

Iteration	SSE	Parameters			
0	1,10237	0,100	0,100	0,090	
1	0,26502	-0,046	0,246	-0,027	
2	0,18848	0,028	0,396	-0,014	
3	0,15119	0,101	0,546	-0,006	
4	0,13435	0,169	0,696	-0,002	
5	0,12828	0,224	0,846	-0,000	
6	0,12799	0,216	0,868	-0,000	
7	0,12796	0,220	0,875	-0,000	
8	0,12795	0,224	0,879	-0,000	
9	0,12795	0,228	0,882	-0,000	
10	0,12795	0,230	0,884	-0,000	
11	0,12795	0,232	0,886	-0,000	

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type		Coef	SE Coef
SAR	4	0,2317	0,1403
SMA	4	0,8855	0,0867
Constant		-0,0001128	0,0005739

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 97, after differencing 92

Residuals: SS = 0,125022 (backforecasts excluded)
MS = 0,001405 DF = 89

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

Lag	12	24	36	48
Chi-Square	7,2	11,3	23,8	36,0
DF	9	21	33	45
P-Value	0,620	0,956	0,880	0,830

Forecasts from period 97

Period	Forecast	95% Limits		Actual
		Lower	Upper	
98	6,51030	6,43683	6,58378	
99	6,53415	6,43024	6,63806	
100	6,53354	6,40628	6,66081	
101	6,49809	6,35114	6,64504	

