



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2012-08-17
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	TER2
Tid	8-12
Kurskod	732G25
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Statistisk analys av samhällsdata Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	5
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel:1463
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. Anteckning på flärp tillåten.
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	17

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2012-08-17

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. Anteckning på flärp tillåten.

Jourhavande lärare: Lotta Hallberg.

Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

1

I tabellen nedan visas totalförsäljningen (i löpande priser) för varugrupperna Husvagnar och Husbilar för ett visst företag, samt priser för husvagnstyp Kabbevagn från varugrupp Husvagnar och husbilstyp Kabbebil från varugrupp Husbilar under åren 2009 till 2011. (alla uppgifter är påhittade)

År	Tot. förs. värde, mkr, husvagnar	Tot. förs. värde, mkr, husbilar	Pris, tkr, Kabbevagn	Pris, tkr, Kabbebil
2009	335	185	250	750
2010	350	190	290	800
2011	370	210	320	830

a) Använd varorna Kabbevagn och Kabbebil som representantvaror för sina varugrupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2009.

2p

b) Hur har företagets priser för husvagnar och husbilar förändrats mellan 2010 och 2011?

1p

Formler finner du längst bak i tentan.

2

I tabellen nedan visas kvinnors folkmängd i åldrarna 14 till 49 år, år 2011 i Östergötland. Även antalet levande födda av dessa kvinnor är angivna. Källa: SCB

Kvinnans ålder	Folkmängd, kvinnor	Antal födda
15 år	2390	3
16 år	2533	3
17 år	2799	10
18 år	2905	25
19 år	3088	33
20 år	3240	53
21 år	3316	107
22 år	3037	113
23 år	3043	155
24 år	2857	191

25 år	2811	225
26 år	2619	274
27 år	2405	256
28 år	2422	324
29 år	2316	318
30 år	2404	324
31 år	2468	382
32 år	2389	325
33 år	2425	322
34 år	2460	269
35 år	2553	243
36 år	2566	207
37 år	2795	176
38 år	2715	132
39 år	2721	98
40 år	2763	93
41 år	2699	40
42 år	2647	33
43 år	2748	18
44 år	2938	9
45 år	3067	4
46 år	3070	4
47 år	2936	1
48 år	2664	0
49 år	2663	0

Slå först ihop folkmängd och antalet levande födda i femårs-intervall (efter kvinnans ålder). Använd därefter dessa data och beräkna den summerade fruktsamheten för antalet födda barn per kvinna i Östergötland år 2011. 3p

3

Man vill analysera en tidsserie som beskriver den kvartalsvisa förbrukningen av naturgas för livsmedelsindustrin under en 10-årsperiod.

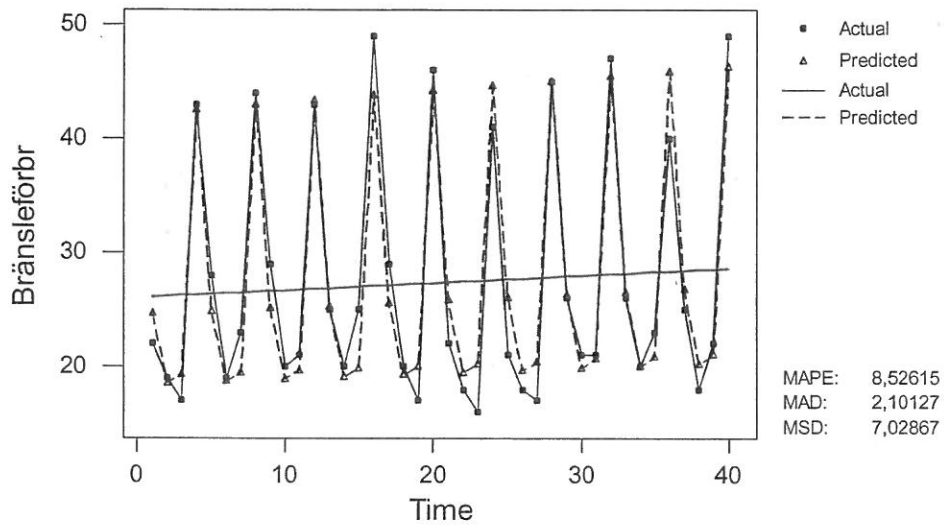
Först betraktar vi tidsseriens tolv första termer som anger förbrukningen i 1000-tal enheter för de tre första åren (12 kvartal) under perioden:

22, 18, 16, 42, 28, 20, 24, 44, 30, 20, 22, 46

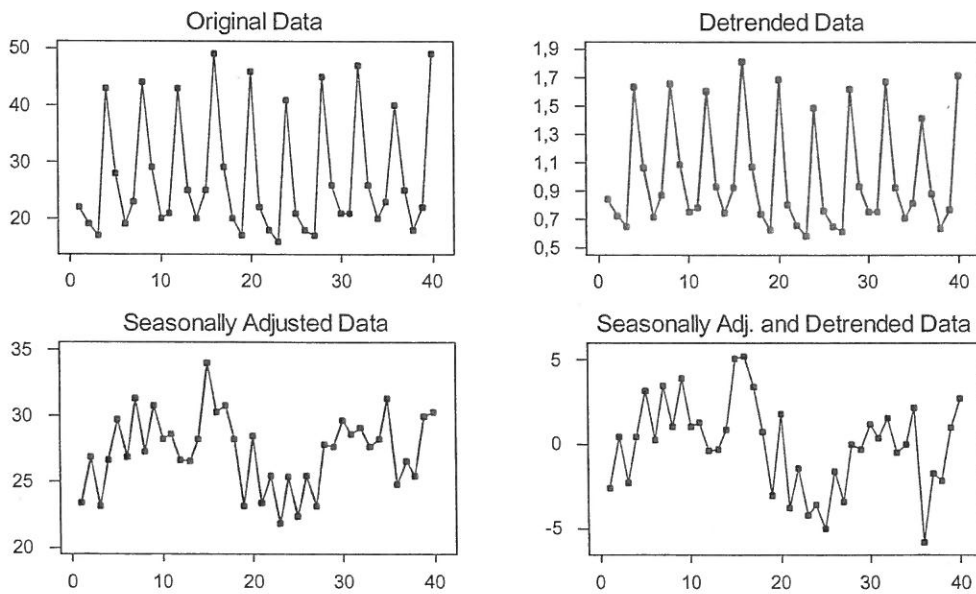
- a) Beskriv serien med 4-punkts centrerade glidande medeltal och skatta därefter de fyra säsongeffekterna enligt additiv metod. Tolka i enkla ordalag de fyra säsongskattningarna. 3p

I fyra bilder följer nu en multiplikativ komponentuppdelning av hela tidsserien (40 termer).

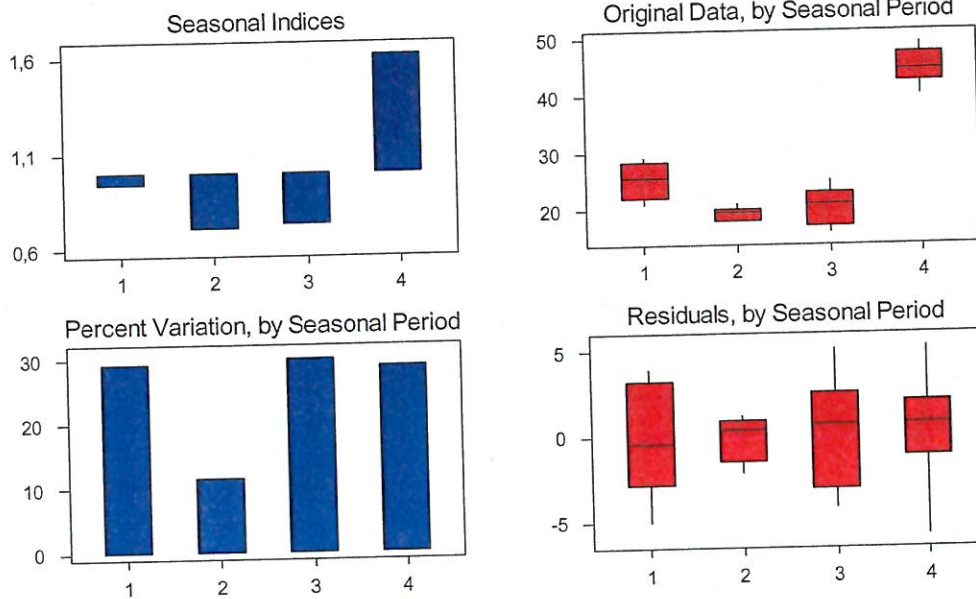
Förbrukning av naturgas, multiplikativ modell



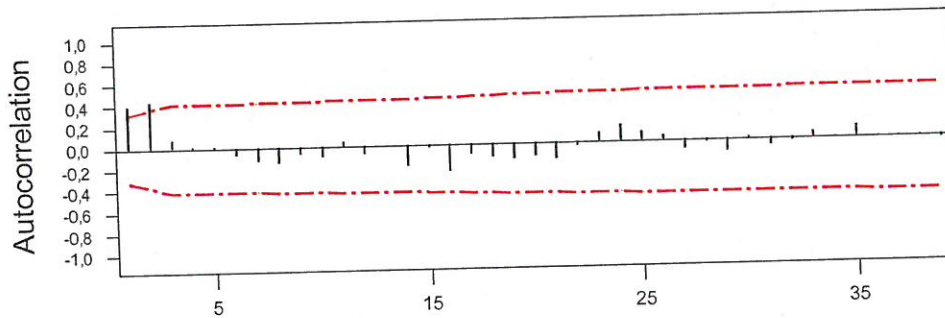
Förbrukning av naturgas, multiplikativ modell



Förbrukning av naturgas, multiplikativ modell



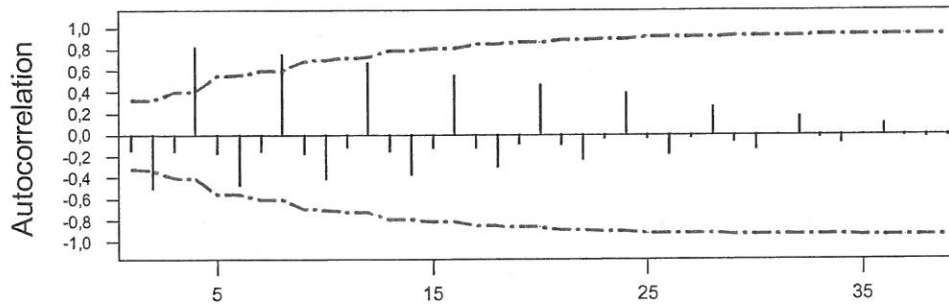
Residualer efter multiplikativ komponentuppdelning



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0,40	2,50	6,73	13	0,01	0,04	18,59	25	0,08	0,34	36,32	37	-0,00	-0,00	44,86
2	0,43	2,40	15,07	14	-0,19	-0,89	20,94	26	0,05	0,19	36,57	38	-0,00	-0,02	44,88
3	0,09	0,43	15,42	15	-0,03	-0,12	20,99	27	-0,09	-0,36	37,54	39	-0,03	-0,10	45,95
4	0,02	0,08	15,43	16	-0,25	-1,15	25,36	28	-0,03	-0,11	37,64				
5	0,01	0,05	15,44	17	-0,09	-0,42	26,01	29	-0,11	-0,48	39,65				
6	-0,06	-0,30	15,62	18	-0,13	-0,57	27,26	30	0,02	0,07	39,70				
7	-0,12	-0,59	16,37	19	-0,14	-0,61	28,79	31	-0,06	-0,27	40,46				
8	-0,14	-0,67	17,41	20	-0,13	-0,56	30,17	32	-0,02	-0,09	40,56				
9	-0,07	-0,33	17,67	21	-0,16	-0,67	32,31	33	0,06	0,24	41,39				
10	-0,09	-0,43	18,14	22	-0,03	-0,13	32,40	34	-0,01	-0,02	41,40				
11	0,05	0,22	18,27	23	0,08	0,34	33,03	35	0,10	0,42	44,83				
12	-0,07	-0,34	18,58	24	0,16	0,67	35,60	36	0,01	0,04	44,86				

Och till slut en bild över den ursprungliga seriens autokorrelationsfunktion SAC:

Originalserien, autocorrelation

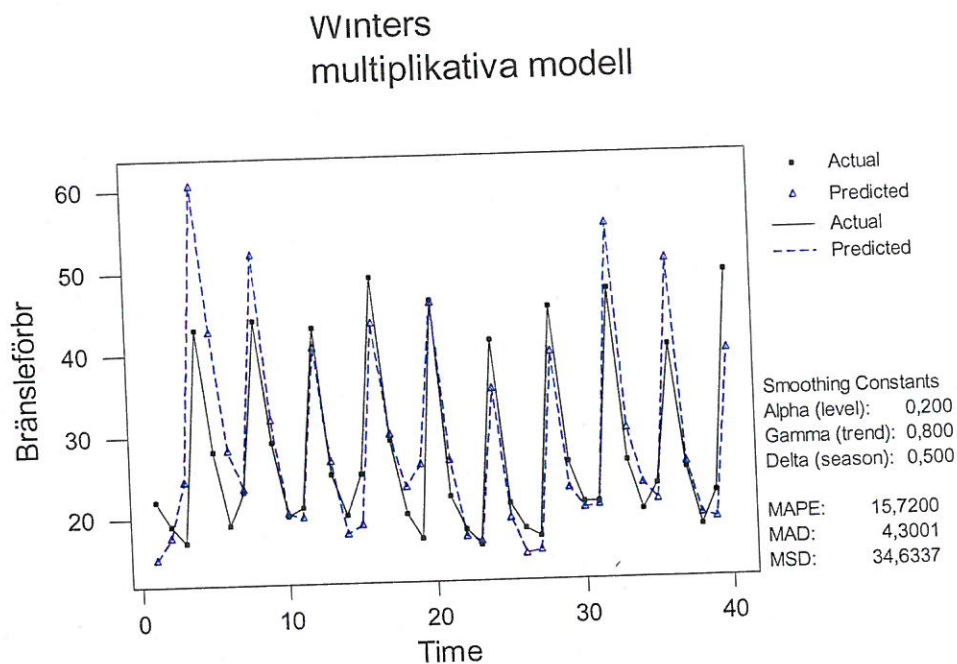
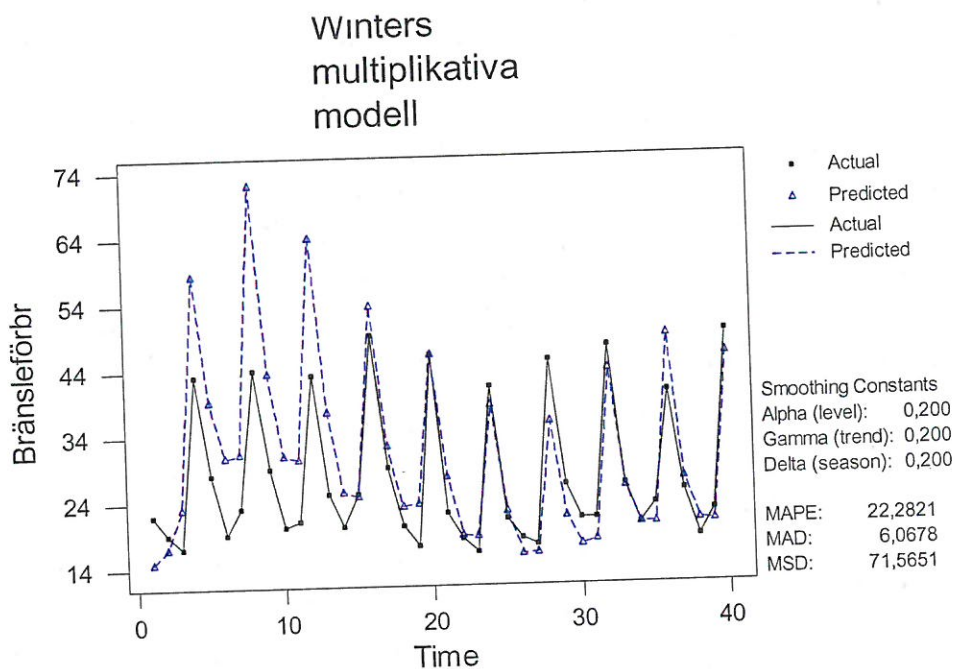


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0,15	-0,95	0,97	13	-0,16	-0,41	130,81	25	-0,05	-0,11	213,73	37	-0,03	-0,06	245,97
2	-0,50	-3,12	12,20	14	-0,38	-0,97	140,16	26	-0,19	-0,42	217,97	38	-0,03	-0,07	246,94
3	-0,17	-0,84	13,44	15	-0,14	-0,34	141,40	27	-0,04	-0,08	218,15	39	-0,03	-0,06	248,08
4	0,82	4,11	45,17	16	0,55	1,38	162,94	28	0,26	0,57	227,57				
5	-0,18	-0,67	46,78	17	-0,14	-0,33	164,31	29	-0,07	-0,15	228,32				
6	-0,48	-1,74	58,18	18	-0,31	-0,74	171,65	30	-0,15	-0,32	231,90				
7	-0,16	-0,56	59,56	19	-0,09	-0,22	172,36	31	-0,02	-0,03	231,95				
8	0,76	2,54	89,51	20	0,47	1,10	191,00	32	0,17	0,38	238,20				
9	-0,19	-0,54	91,38	21	-0,10	-0,23	191,91	33	-0,03	-0,07	238,46				
10	-0,42	-1,21	101,21	22	-0,25	-0,56	197,52	34	-0,09	-0,19	240,57				
11	-0,13	-0,36	102,14	23	-0,04	-0,10	197,70	35	-0,00	-0,01	240,58				
12	0,67	1,87	129,20	24	0,39	0,87	213,48	36	0,11	0,23	245,51				

- b) Beskriv i enkla ordalag resultatet av komponentuppdelningen med avseende på trend, cyklisk variation, säsongvariation och residualer. Följande frågor skall besvaras med kort motivering och tolkning:
- Ökar/minskar förbrukningen under perioden?
 - Verkar det finnas cyklisk variation?
 - Hur varierar säsongerna?
 - Verkar residualerna vara "vitt brus" (slumpmässiga)?
 - Bör residualerna modelleras? Med vilken modell?
 - Kan man på det hela taget tycka att komponentuppdelningen varit lyckad? 3p

4

Vi utnyttjar samma tidsserie som i föregående uppgift.
Vi analyserar tidsserien enligt Holt-Winters modell med två olika uppsättningar utjämningskonstanter. För den första är konstanterna 0.2, 0.2, 0.2 och för den andra 0.2, 0.8, 0.5.



a) Vilken av de båda modellerna föredrar du? Motivera ditt val. Försök också att kort förklara varför den modell som Du valt ger bättre resultat än den andra. 1p

Vi ska nu göra prognoser för det kommande året.

Vi börjar med att utnyttja en enkel linjär regressionsanalys där vi låter tidsserien (Y_t), "förklaras" av tiden (t : 1, 2, 3, ..., 39, 40). Regressionsanalysen redovisas nedan:

Regression Analysis: Y_t versus tidpunkt

The regression equation is
 $Y_t = 26,1 + 0,064 \text{ tidpunkt}$

Predictor	Coef	SE Coef	T	P
Constant	26,062	3,489	7,47	0,000
tidpunkt	0,0641	0,1483	0,43	0,668

S = 10,83 R-Sq = 0,5

b) Beräkna prognoser för de fyra följande kvartalen.

1p

Vi utnyttjar nu i stället Holt-Winters multiplikativa metod med konstanterna 0.2, 0.8, 0.5 anpassad till hela tidsserien om 40 observationer för att göra prognoser. I tabellen redovisas resultatet för de fem sista kvartalen i serien (kvartalen 36, 37, 38, 39 och 40).

tidpunkt	Y_t	SMOO	LEVE	TREN	SEAS	FITS
36	40	50,2528	31,6696	-0,9405	1,39559	50,495
37	25	26,3505	30,5926	-1,0497	0,82462	25,568
38	18	19,9534	29,1538	-1,361	0,63482	19,269
39	22	19,7163	28,7404	-0,6029	0,72088	18,796
40	49	40,1098	29,5321	0,51277	1,5274	39,268

c) Beräkna prognoser för de två följande kvartalen. Förklara också kortfattat varför prognoserna enligt Holt-Winters metod skiljer sig från dem som erhöles under b) ovan.

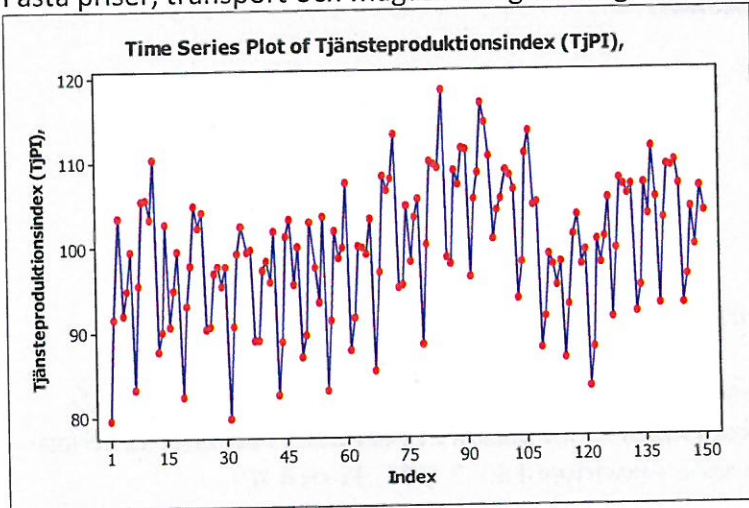
2p

5

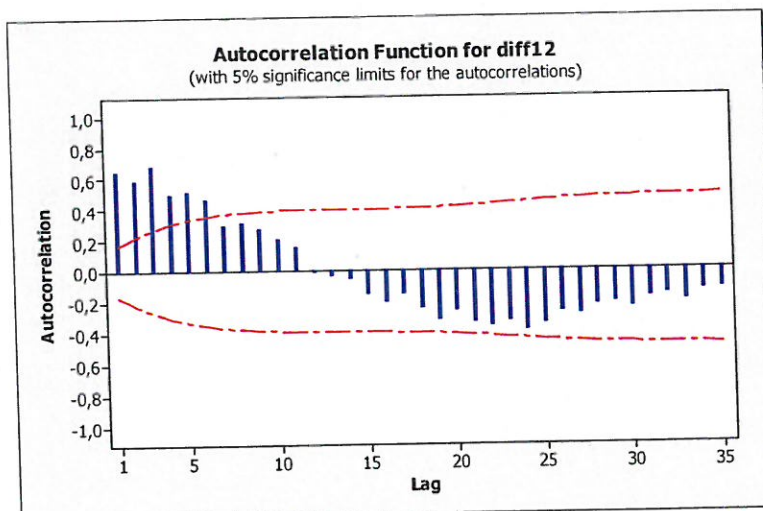
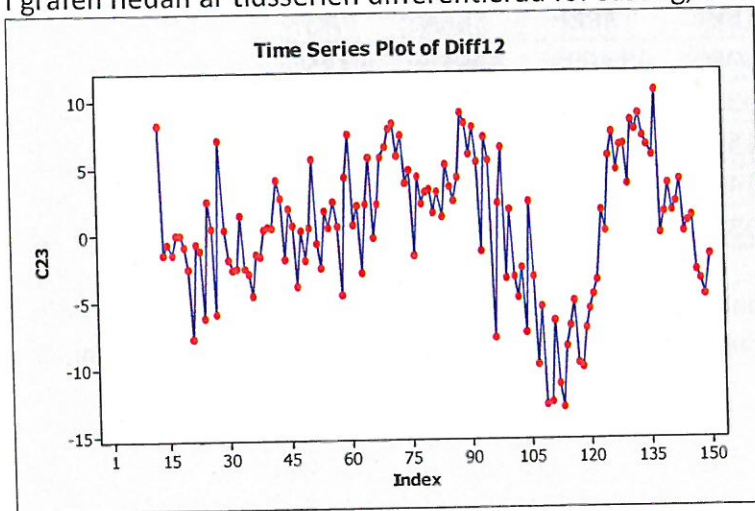
I grafen nedan ses tidsserien:

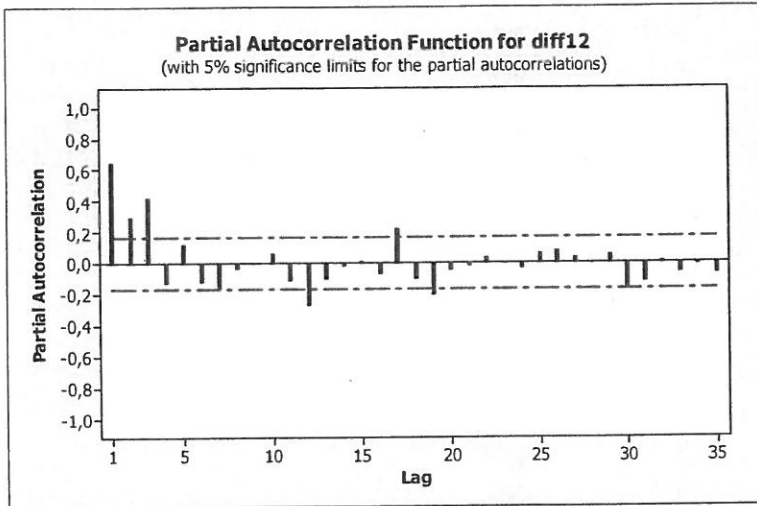
Tjänsteproduktionsindex (TjPI), 2005=100, efter näringsgren SNI 2007. Månad 2000M01-2012M06

Fasta priser, transport och magasineringsföretag. Källa; SCB



I grafen nedan är tidsserien differentierad för säsong, Diff12.





ARIMA Model: Tjänsteproduktionsindex (TjPI),

Estimates at each iteration

Iteration	SSE	Parameters		
0	2826,98	0,100	0,100	0,490
1	2214,12	0,250	0,186	0,331
2	1933,47	0,400	0,269	0,180
3	1911,00	0,454	0,295	0,132
4	1910,82	0,459	0,296	0,137
5	1910,82	0,459	0,296	0,139
6	1910,82	0,459	0,296	0,139
7	1910,82	0,459	0,296	0,139

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,4595	0,0823	5,58	0,000
AR 2	0,2962	0,0825	3,59	0,000
Constant	0,1390	0,3197	0,43	0,664

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 150, after differencing 138

Residuals: SS = 1903,73 (backforecasts excluded)
 MS = 14,10 DF = 135

Nedan ges de 13 sista värdena i tiddserien.

Tjänsteproduktionsindex (TjPI),	Diff12	tid
105,5	-0,1	138
92,9	1,5	139
103,1	3,6	140
109,4	1,6	141
109,3	2,2	142
109,9	3,9	143
107,1	0,0	144
92,9	0,8	145
96,3	1,2	146
104,3	-2,9	147
99,9	-3,6	148
106,8	-4,7	149
103,8	-1,7	150

- a) Vilken SARIMA (säsong ARIMA) modell är mest lämplig för dessa data. Använd graferna ovan för att bestämma lämplig modell. 2p
- b) Beräkna en prognos för de två nästkommande tidpunkterna med den anpassade modellen i utskriften ovan. 2p

Formelsamling Index

Sammansatta fastbasindex:

$$I_t = i_{1,t} \cdot w_1 + i_{2,t} \cdot w_2 + \dots + i_{n,t} \cdot w_n$$

där n är antalet ingående varor/tjänster, $i_{1,t}, \dots, i_{n,t}$ är enkla prisindex för ingående varor, alla med basår t_0 och w_1, \dots, w_n väljs enligt ett viktsystem:

$$\text{Laspeyre: } w_i = \frac{p_{i,t_0} \cdot q_{i,t_0}}{\sum_j p_{j,t_0} \cdot q_{j,t_0}}$$

$$\text{Paasche: } w_i = \frac{p_{i,t_0} \cdot q_{i,t}}{\sum_j p_{j,t_0} \cdot q_{j,t}}$$

Kedjeprisindex:

$$I_t = \frac{L_{0,1}}{100} \cdot \frac{L_{1,2}}{100} \cdot \dots \cdot \frac{L_{t-1,t}}{100} \cdot 100$$

där

$$L_{t-1,t} = \sum_{i=1}^n \frac{p_{i,t}}{p_{i,t-1}} \cdot 100 \cdot w_{i,t-1,t}$$

är årslänken från år $t-1$ till t för n ingående varor/tjänster. $w_{i,t-1,t}$ väljs enligt ett viktsystem:

$$\text{Laspeyre: } w_{i,t-1,t}^L = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t-1}{\text{Totala försäljningsvärdet år } t-1}$$

$$\text{Paasche: } w_{i,t-1,t}^P = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t \text{ i priser för år } t-1}{\text{Totala försäljningsvärdet år } t \text{ i priser för år } t-1}$$

Med representantvaror byts "Försäljningsvärdet för vara i " mot "Försäljningsvärdet för varugrupp i " i vikterna.

Implicitprisindex:

$$I_t = \frac{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i löpande priser}}{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i basårets priser}} \cdot 100$$

Relativprisindex:

$$I_t^R = \frac{I_t^v}{I_t^0} \cdot 100$$

där I_t^v = Prisindex för aktuell vara/tjänst/grupp och I_t^0 = Prisindex för den större jämförelsegruppen, t ex KPI.