



# Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2013-10-19
<b>Sal (2)</b> Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER1 TERD
<b>Tid</b>	8-12
<b>Kurskod</b>	732G25
<b>Provkod</b>	TEN1
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Statistisk analys av samhällsdata Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	4
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Lotta Hallberg
<b>Telefon under skrivtiden</b>	
<b>Besöker salen ca kl.</b>	10
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	annelie.almquist@liu.se, tel 2934
<b>Tillåtna hjälpmedel</b>	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. Liten anteckning på flärp tillåten.
<b>Övrigt</b>	
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	

# TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2013-10-19

**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. Liten 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 ( värdeserie i löpande priser ) för varugrupperna A och B för ett visst företag, samt priser för varan A1 från varugrupp A och varan B1 från varugrupp B1 under åren 2010 till 2012

År	Tot. förs. A	Tot. förs. B	Pris, A1	Pris, B1
2010	335	185	23	16
2011	350	190	24	16
2012	370	210	25	17

- a) Använd varorna A1 och B1 som representantvaror för sina varugrupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2010. 2p
- b) Hur har företagets priser förändrats mellan 2010 och 2012? Svara med tolkning av ett tal. 1p

2

Följande livslängdstabell är för kvinnor i ett visst land år 2010-2012.

Ålder $x$	Antal överlevande till ålder $x$ , $l_x$	Totala antalet kvinnoår levda över ålder $x$ , $T_x$
0	100 000	7 700 187
1	99 016	7 601 014
10	98 746	6 711 410
20	98 497	5 725 014
30	98 105	4 741 877
40	97 346	3 764 073

- a) Beräkna sannolikheten att en kvinna som är 20 år överlever sin 40 årsdag. 2p
- b) Beräkna spädbarnsdödligheten. 1p
- c) Beräkna den återstående förväntade livslängden för en flicka som överlevt sitt första levnadsår. 1p
- d) Beräkna den förväntade livslängden för en kvinna som uppnått en ålder på 20 år. 1p

**Formler:**

$${}_n q_x = \frac{l_x - l_{x+n}}{l_x} \quad {}_n L_x = \frac{n(l_x + l_{x+n})}{2}$$

$$e_x = \frac{T_x}{l_x} \quad T_x = \sum_{i=x} {}_n L_i$$

### 3

Detta exempel är baserat på ett exempel i Vejde-Leander, Ordbok i statistik.

En sommarförsäljare av glass är intresserad av att göra bra prognoser för hur mycket glass som hon kommer att sälja olika dagar. Hon använder de första sex dagarna under försäljningsperioden som "testperiod". Hon noterar varje dag morgontemperaturen i grader och försäld mängd glass under dagen i hela liter. I tabellen återfinns data för fem dagar, dagarna 2 till 6. Försäljningen den första dagen var 42 liter. Morgontemperaturen den sjunde dagen var 22 grader.

Tabellen visar såld glassmängd och morgontemperatur för fem dagar i maj.

Dag, nummer	Temp, x	Glassmängd, y
2	14	48
3	12	35
4	18	60
5	16	56
6	25	71

Hon vill pröva två relativt enkla prognosmetoder, nämligen

#### Prognosmetod 1

Hon prognosticerar dagens försäljning med gårdagens.

#### Prognosmetod 2

Hon använder data som beskriver morgontemperaturen en viss dag och försäld mängd samma dag för att med hjälp av regressionsanalys skatta sambandet. Detta samband använder hon sedan för att utifrån morgontemperaturen varje dag skatta hur mycket hon kommer att sälja respektive dag.

Regressions sambandet framtaget med hjälp av Minitab ser ut som följer:

### Regression Analysis: glass versus temp

The regression equation is  $\text{glass} = 13,7 + 2,37 \text{ temp}$

Predictor	Coef	SE Coef	T	P
Constant	13,71	13,01	1,05	0,369
temp	2,3700	0,7401	3,20	0,049

S = 7,401      R-Sq = 77,4%

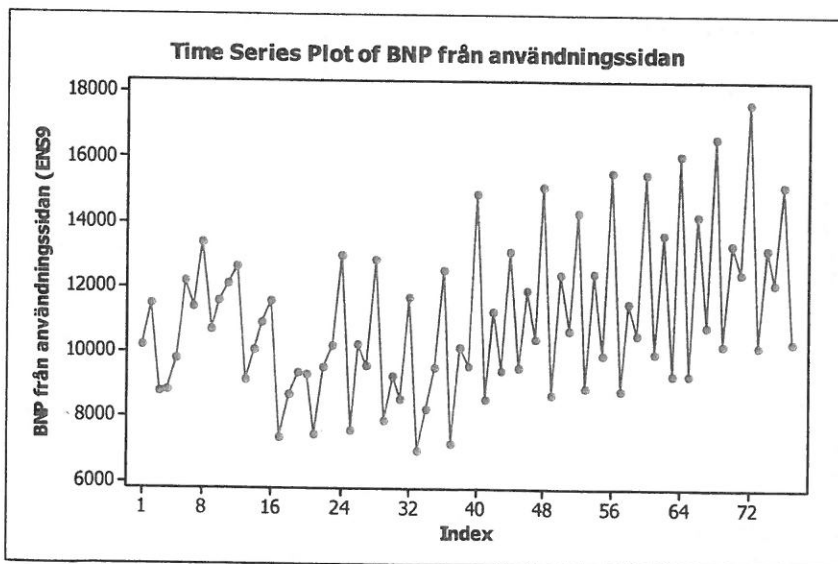
- a) Beräkna de olika måtten för prognosfelet (MAD, MSE/MSD, MAPE) baserat på de fem dagarna 2 – 6 enligt såväl prognosmetod 1 som prognosmetod 2.      3p
- c) Beräkna hur mycket glass försäljaren kan räkna med att sälja dag 7 med var och en av de två metoderna. Jämför också de båda metoderna. Vilken skulle Du välja för att prognosticera just dag 7 och varför? Fundera också på hur det skulle te sig på längre sikt. Glassförsäljaren vill ju göra prognos för varje dag för att kunna bedöma hur mycket glass hon skall ta med. och diskutera vilken som verkar vara rimligast. Diskutera!      2p

## 4

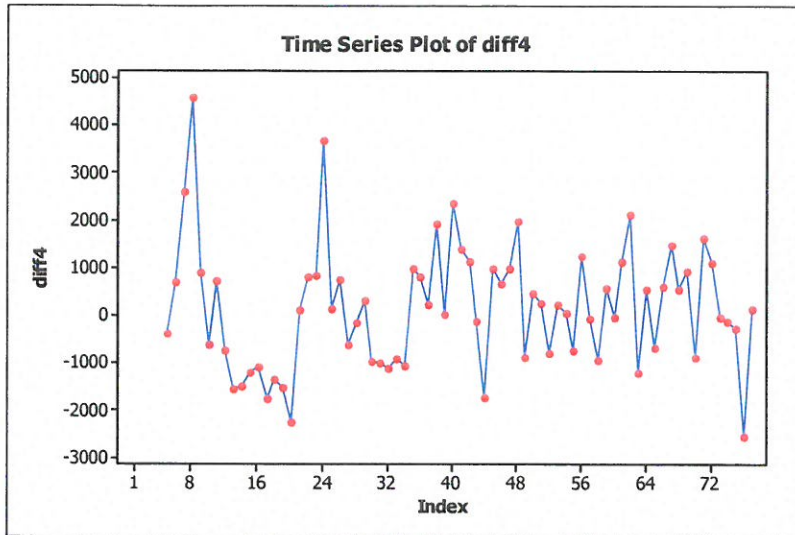
Följande tidsserie är hämtad från SCBs hemsida. Data är:

*BNP från användningssidan, försörjningsbalans, fasta priser referensår 2011, mnkr Statliga myndigheters fasta bruttoinvesteringar.*

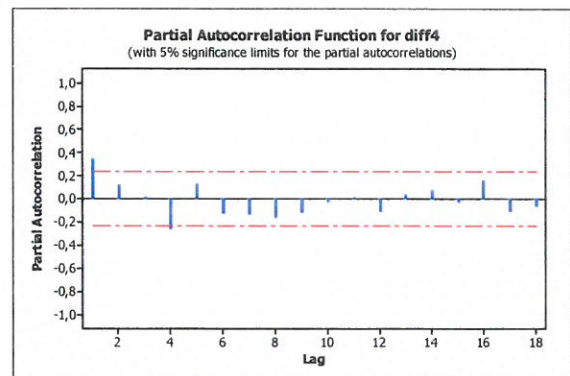
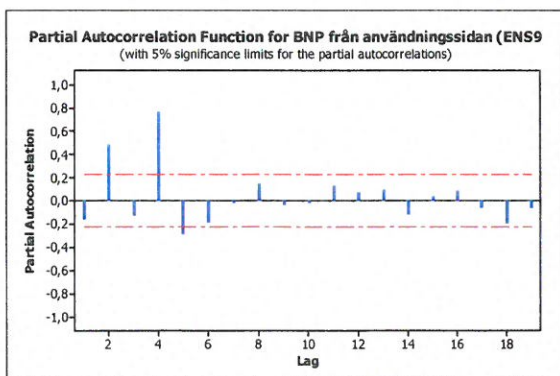
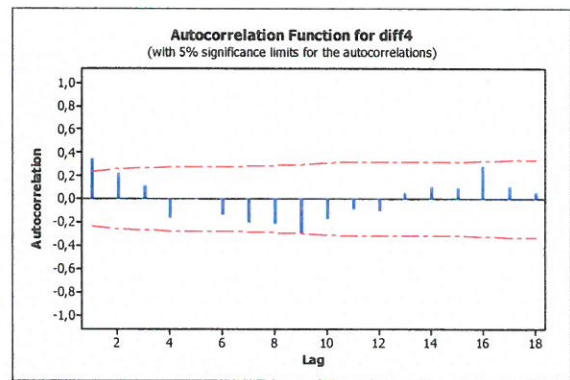
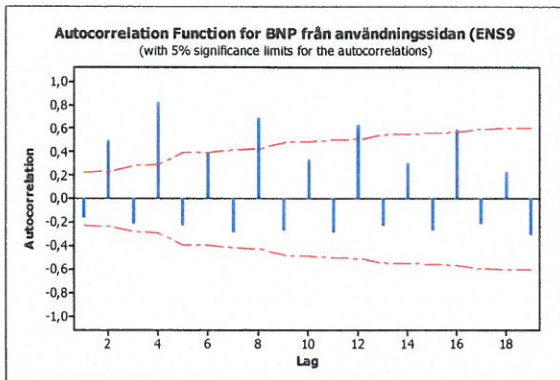
Data är mellan 1993K1 och 2012K1. Kvartalsdata. Vi vill nu bland annat beräkna prognoser för kvartal 2 och 3 år 2012.



Nedan ses den differentierade serien för kvartal



Nedan följer några grafer, SAC och SPAC på originalserie och diff4. Därefter 6 ARIMA/SARIMA modeller



### MODELL 1

#### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0,0903	0,1139	-0,79	0,430
SMA 4	-0,6688	0,0862	-7,76	0,000
Constant	12018,4	317,8	37,82	0,000
Mean	11022,7	291,5		

Number of observations: 77

Residuals: SS = 210351119 (backforecasts excluded)  
MS = 2842583 DF = 74

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

Lag	12	24	36	48
Chi-Square	85,7	158,1	198,8	240,8
DF	9	21	33	45
P-Value	0,000	0,000	0,000	0,000

### MODELL 2

#### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	0,0520	0,1144	0,45	0,651
SMA 4	-0,6685	0,0864	-7,74	0,000
Constant	11022,7	303,5	36,32	0,000
Mean	11022,7	303,5		

Number of observations: 77

Residuals: SS = 211428192 (backforecasts excluded)  
MS = 2857138 DF = 74

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

Lag	12	24	36	48
Chi-Square	93,9	174,3	218,8	261,4
DF	9	21	33	45
P-Value	0,000	0,000	0,000	0,000

### MODELL 3

#### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0,3137	0,1137	-2,76	0,007
SMA 4	0,4461	0,1086	4,11	0,000
Constant	140,8	100,0	1,41	0,164

Differencing: 0 regular, 1 seasonal of order 4

Number of observations: Original series 77, after differencing 73  
Residuals: SS = 94118610 (backforecasts excluded)  
MS = 1344552 DF = 70

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

Lag	12	24	36	48
Chi-Square	13,4	26,7	36,2	43,2
DF	9	21	33	45
P-Value	0,144	0,182	0,322	0,549

## MODELL 4

### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,6599	0,2303	2,87	0,006
SAR 4	0,0377	0,2522	0,15	0,882
MA 1	0,3067	0,2864	1,07	0,288
SMA 4	0,5334	0,2095	2,55	0,013
Constant	42,98	42,76	1,01	0,318

Differencing: 0 regular, 1 seasonal of order 4

Number of observations: Original series 77, after differencing 73

Residuals: SS = 86525201 (backforecasts excluded)  
MS = 1272429 DF = 68

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

Lag	12	24	36	48
Chi-Square	5,6	20,4	29,2	35,5
DF	7	19	31	43
P-Value	0,583	0,370	0,558	0,785

## MODELL 5

### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,4294	0,1089	3,94	0,000
SAR 4	-0,3673	0,1166	-3,15	0,002
Constant	111,0	136,1	0,82	0,417

Differencing: 0 regular, 1 seasonal of order 4

Number of observations: Original series 77, after differencing 73

Residuals: SS = 94614204 (backforecasts excluded)  
MS = 1351631 DF = 70

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

Lag	12	24	36	48
Chi-Square	9,8	23,2	32,3	40,6
DF	9	21	33	45
P-Value	0,363	0,333	0,503	0,657

## MODELL 6

### ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,5450	0,2830	1,93	0,058
MA 1	0,2242	0,3283	0,68	0,497
Constant	67,9	111,9	0,61	0,546

Differencing: 0 regular, 1 seasonal of order 4

Number of observations: Original series 77, after differencing 73

Residuals: SS = 105986406 (backforecasts excluded)

MS = 1514092 DF = 70

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

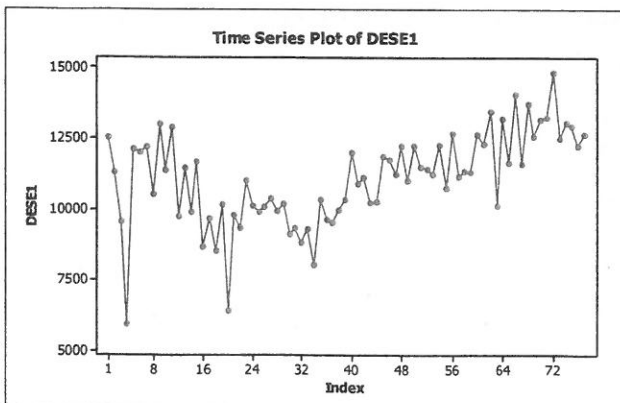
Lag	12	24	36	48
Chi-Square	14,6	36,1	52,8	62,0
DF	9	21	33	45
P-Value	0,104	0,021	0,016	0,047

BNP	diff4	Resid	tid
13216	-129	-302,33	74
12176	-272	-337,34	75
15147	-2563	-2558,25	76
10324	126	881,46	77

- Seriens värden är i fasta priser med referensår 2011. Vad innebär det? 1p
- Ovan har sex ARIMA/SARIMA modeller anpassats. Ta hjälp av relevanta utskrifter och välj den modell som är bäst. Motivera väl. 2p
- Gör prognos för kvartal 2 och 3 år 2012 genom att använda modell 6. 2p

I nästa modell (som vi kallar modell 7) så har först serien säsongrensats med klassisk additiv komponentuppdelning. Säsongrensad serie kallas här DESE1.

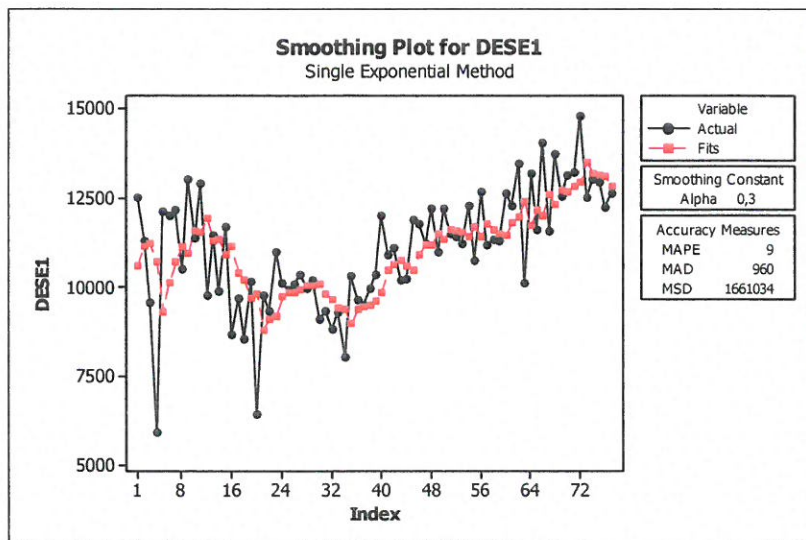
Sedan har enkel exponentiell utjämning utförts på DESE1.





### Säsongskomponenter

Period	Index
1	-2309,02
2	184,55
3	-772,64
4	2897,11



DESE1	utjämnad serie $l_t$	t
13031,5	13152,7	74
12948,6	13091,5	75
12249,9	12839,0	76
12633,0	12777,2	77

- d) Gör prognos för BNP för kvartal 2 och 3 år 2012 genom att använda modell 7. 2p
- e) Tolka säsongskomponenten för kvartal 2. 1p
- f) Visa hur  $l_{76}$  och  $l_{77}$  har beräknats. (Utgå från att  $l_{75}$  är beräknad) 1p