2004/29

Notater

2004

Dinh Quang Pham

Sesongjustering for boligprisindeksen

Notater

Seksjon for Metoder og standarder

Statistics Norway

Statistisk sentralbyrå

Innhold

Figurer

1 Innledning

Seksjon for bygg og tjenestestatistikk (S460) har nettopp utviklet en ny metode for å beregne kvartalsvis prisindeks for boligmarkedet fra og med første kvartal 2003. Indeksene beregnes på grunnlag av data fra Finn.no og Norske Boligbyggelags Landsforbund (NBBL). Tidligere hentet vi opplysninger om kjøp og salg av bruktboliger fra Tinglysningsregisteret kombinert med en skjemaundersøkelse til aktuelle boligkjøpere. Fordelene er

- Forbedret aktualitet: Tallene blir publisert raskere.

- Med data fra Finn.no og NBBL er det ikke lenger nøvendig å publisere foreløpige tall.

- Lavere kosnader: Siden vi ikke bruker skjemaundersøkelsen lenger.

Datagrunnlag og beregningsmetode for boligprisindeksen er nærmere beskrevet i Notat 2003/83 av Thor Herman Christensen. Seksjon for statistiske metoder og standarder skal i samarbeid med S460 vurdere bruk av sesongjustering for 5 tidsseriene. Det er

1. Oslo og Bærum (vi skriver Oslo-B for enkelthets skyld),
2. Akershus utenom Bærum (Akershus),
3. Stavanger, Bergen og Trondheim (Storby),
4. Resten av landet (Resten).
5. I alt (I alt).

Tidsseriene er observert fra første kvartal 1992 til fjerde kvartal 2003. I alt er det 48 observasjoner i hver av tidsseriene. Vi bruker X12-ARIMA for sesongjusteringen. Resultatene er presentert i de neste avsnittene.

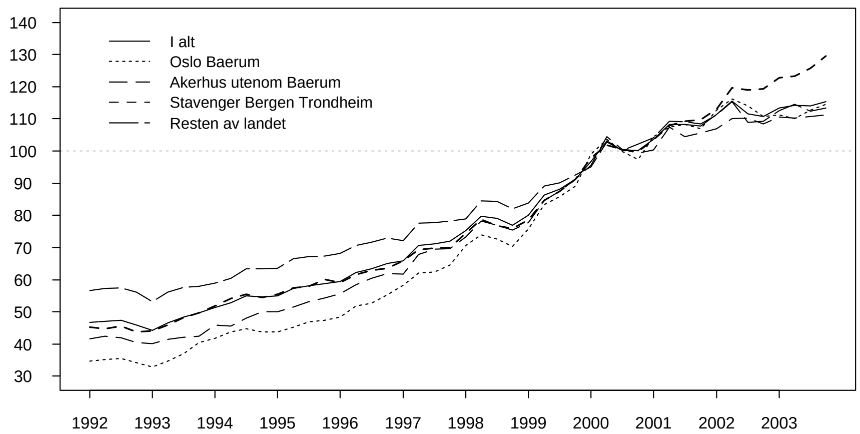
2 Data

Data for de 5 tidsseriene med basisår 2000, er plottet i figur 1. Prisene økte over hele landet. For å kunne se endringene i hele tidsserien siden 1992, beregner vi tallene med basisår 1992. La Y_t være prisindeksen i tidspunkt t, med basisår 2000. Z_t er prisindeksen med basisår 1992. Vi beregner Z_t fra Y_t ved

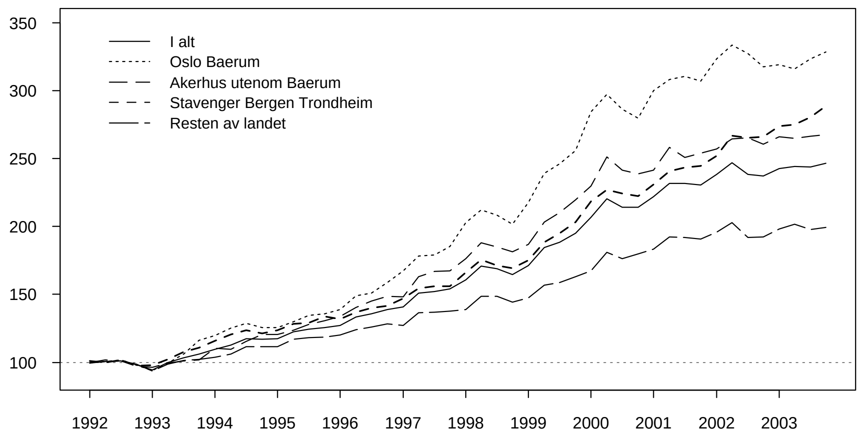
$$Z_t = 100 \times Y_t/\overline{Y}_{1992}$$

der $\overline{Y}_{1992}$ er gjennomsnitt over kvartalene i 1992. Z_t, er plottet i figur 2. Tabell 1 og 2 viser tallene med basisår 2000 og 1992. Vi ser at prisene økte over hele landet og mest i Stavanger, Bergen og Trondheim siden 2000, men i hele tidsserien tilbake til 1992 er det i Oslo og Bærum prisene har økt mest. Endringen fra 1.kvartal 1992 til 4.kvartal 2003 for Oslo-Bærum er 328,6-99,6=229,0, og for Stavanger, Bergen, Trondheim er 288,7-101,0=187,7 (se tabell 2). Relative endringen fra tidspunkt $t-1$ til t ved basisår 2000 eller 1992 er det samme.

Vi presenterer resultater fra X12-ARIMA for 5 tidsserier ved å benytte basisår lik 2000.



Figur 1: *Prisindeksene med basisår 2000.*



Figur 2: *Prisindeksene med basisår 1992.*

Tabell 1: *Basisår 2000. Prisindeksene.*

	I alt	Oslo-B	Akershus	Stavnger-	Resten
1kv.1992	46.7	34.7	41.6	45.3	56.6
.	.	.	.	.	.
3kv.2003	114.0	112.7	110.8	125.8	112.4
4kv.2003	115.3	114.5	111.3	129.5	113.3

Tabell 2: *Basisår 1992. Prisindeksene.*

	I alt	Oslo-B	Akershus	Stavnger-	Resten
1kv.1992	99.9	99.6	100.0	101.0	99.6
.	.	.	.	.	.
3kv.2003	243.9	323.4	266.3	280.5	197.8
4kv.2003	246.6	328.6	267.5	288.7	199.4

Vi betegner:

Y_t^{Ialt}	tidsserien for i alt.
Y_t^{Oslo-B}	tidsserien for Oslo og Bærum
$Y_t^{Akershus}$	tidsserien for Akershus utenom Bærum
Y_t^{Storby}	tidsserien for Stavanger, Bergen og Trondheim
Y_t^{Resten}	tidsserien for resten av landet

3 Resultater

Tallene går fra med 1.kvartal 1992 til og med 4.kvartal 2003. I alt er det 48 observasjoner. Den multiplikative modellen er brukt i sesongjusteringen. Vi bruker symbolene S_t, A_t, T_t og I_t som står for sesongkomponenten, sesongjusterte tall, trend og den irregulære komponenten. Vi får følgende resultater

- Rådata

De relative endringene i prosent fra kvartal til kvartal av rådata, $(100\% \times (Y_t - Y_{t-1})/Y_{t-1})$, er vist i tabell 3 og 4. Et generelt trekk er at endringene er store fra første til andre kvartal. Vi plotter rådata av 4 kvartaler av tidsserien "I alt" for å illustrere endringene i tabell 3 og 4.

Tabell 3: *Relative endringer i prosent av rådata for i alt, Oslo-Bærum og Akershus utenom Bærum*

	I alt				Oslo-Bærum				Akershus			
	1.kv	2.kv	3.kv	4.kv	1.kv	2.kv	3.kv	4.kv	1.kv	2.kv	3.kv	4.kv
1992		0.9	0.4	-3.0		1.2	1.1	-3.9		1.9	-0.9	-3.8
1993	-3.7	5.4	3.6	2.9	-3.8	5.5	6.9	9.5	-0.7	3.2	1.7	0.7
1994	3.2	2.9	4.0	-0.4	3.2	4.5	2.5	-2.2	8.3	-0.7	5.3	4.4
1995	0.4	4.2	1.6	1.2	0.0	3.4	3.5	0.9	0.0	2.8	3.3	2.1
1996	1.2	4.7	1.8	2.5	2.3	7.2	1.3	5.1	2.6	5.0	3.2	2.5
1997	1.4	7.1	0.8	1.1	5.4	6.5	0.5	3.5	-0.3	10.0	2.4	0.1
1998	4.4	6.1	-1.0	-2.7	9.4	4.5	-1.8	-3.2	5.3	6.8	-1.8	-2.0
1999	4.0	7.9	2.1	3.5	7.8	10.0	2.9	4.0	3.1	8.9	3.3	4.5
2000	6.0	6.6	-2.9	0.0	11.1	4.5	-3.7	-2.4	4.6	9.4	-3.8	-1.1
2001	3.6	4.4	0.0	-0.5	7.3	2.8	0.7	-1.1	1.0	7.0	-2.8	1.2
2002	3.3	3.7	-3.5	-0.6	5.3	3.1	-1.9	-2.9	1.1	2.9	0.3	-1.7
2003	2.3	0.7	-0.2	1.1	0.5	-1.0	2.4	1.6	2.0	-0.4	0.5	0.5

- ARIMA modell

ARIMA modellene for de 5 tidsseriene er listet ut i tabell 5. Den enkleste modellen velges om følgende 3 kriterier er oppfylt.

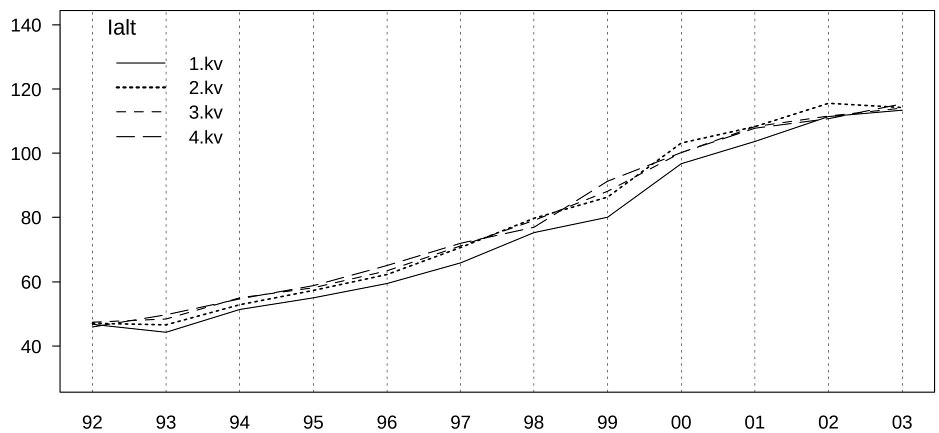
(i) Gjennomsnittlig absolutt prognosefeil i prosent siste 3 år er mindre eller lik 15%.

(ii) Kji-kvadrat (χ^2) for testing av om residualene er ukorrelerte må ha p-verdi større enn 5%.

(iii) Det er ingen tegn til "overdifferencing"(i.e, $\sum_{i=1}^{q} \theta_i > 0.9$, eller $\sum_{i=1}^{Q} \Theta_i > 0.9$, se side 51 i X12-ARIMA manualen).

Tabell 4: *Relative endringer i prosent av rådata for Stavanger-Bergen-Trondheim og resten av landet*

	Storby				Resten			
	1.kv	2.kv	3.kv	4.kv	1.kv	2.kv	3.kv	4.kv
1992		-1.1	1.8	-4.2		1.1	0.3	-2.3
1993	0.7	4.3	5.0	3.1	-5.2	5.6	2.5	0.7
1994	4.4	4.2	2.4	-1.8	1.6	2.5	5.0	0.0
1995	2.0	3.6	0.9	3.6	0.2	4.9	0.8	0.4
1996	-1.7	4.1	2.3	1.1	1.2	3.5	1.6	1.7
1997	3.6	5.2	0.9	0.1	-1.0	7.3	0.4	0.5
1998	6.4	5.6	-2.4	-1.2	0.9	7.1	-0.1	-2.7
1999	3.6	7.6	3.3	4.3	2.2	6.2	1.1	2.8
2000	7.3	4.1	-1.4	-0.7	2.6	8.2	-2.6	2.0
2001	3.7	4.3	1.2	0.5	2.0	4.9	-0.1	-0.6
2002	3.0	5.8	-0.5	0.3	2.7	3.6	-5.5	0.2
2003	2.8	0.4	2.0	2.9	3.1	1.7	-1.8	0.8



Figur 3: *I alt. Rådata for 4 kvartaler*

Tabell 5: *ARIMA modell*

	modell	$\widehat{\theta}$	$std(\widehat{\theta})$	$\widehat{\Theta}$	$std(\widehat{\Theta})$
I alt	(0 1 1)(0 1 1)	-0.3735	0.1392	0.5820	0.1221
Oslo-B	(0 1 1)(0 1 1)	-0.5218	0.1291	0.5429	0.1286
Akershus	(0 1 1)(0 1 1)	-0.0310	0.1513	0.6476	0.1259
Storby	(0 1 1)(0 1 1)	-0.2354	0.1326	0.7823	0.0899
Resten	(0 1 1)(0 1 1)	-0.2354	0.1326	0.7823	0.0899

$Y_t \sim$ARIMA(0 1 1)(0 1 1) med kvartalstall. Ligningen for Y_t er

$$Y_t = Y_{t-1} + Y_{t-4} - Y_{t-5} + \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-4} + \theta\Theta\epsilon_{t-5} \tag{1}$$

Ligning (1) brukes for å forlenge tidsserien og estimere effektene av ukedager, påske og intervensjoner.

- Ekstremverdier

Det er ikke ekstremverdier i noen av tidsseriene.

- Tester for sesongmønster og sesongbevegelser

Testene er vist i tabell 6. Det er sesongvariasjoner i data, men det er ingen tegn til sesongbevegelser i tidsseriene. Sesongjusteringen blir dårlig når sesongvariasjonene er store eller de endrer seg med tiden.

Tabell 6: *Tester for sesongmønster og sesongbevegelser*

	sesongmønster		sesongbevegelse	
I alt	F=21.37	sign. på 5%	F=0.57	ikke sign. på 5%
Oslo-B	F=11.40	sign. på 5%	F=1.13	ikke sign. på 5%
Akershus	F=11.68	sign. på 5%	F=1.86	ikke sign. på 5%
Storby	F=19.68	sign. på 5%	F=0.48	ikke sign. på 5%
Resten	F=19.70	sign. på 5%	F=0.69	ikke sign. på 5%

- Framskrivinger

Det er en økning i trenden. Framskrivingene av rådata, sesongvariasjoner og sesongjusterte tall for 4 kvartaler i 2004 er listet ut i tabell 7 og 8. Boligprisnivået fortsatt øker til neste år. Sesongmønsteret for prisene i boligmarkedet i 2003 og 2004 er det samme. Det er andre kvartal som er høyest og fjerde kvartal som er lavest. Forskjellen mellom det andre kvartalet og de andre er ikke så stor. For eksempel, for "I alt" er $S_{2kv,2004} - S_{1kv,2004} = 102,1 - 100,3 = 1,8\%$, $S_{2kv,2004} - S_{3kv,2004} = 102,1 - 99,3 = 2,8\%$, og $S_{2kv,2004} - S_{4kv,2004} = 102,1 - 98,5 = 3,6\%$. De sesongjusterte tallene i 2004 er også høyere enn i de samme kvartalene i 2003.

Tabell 7: *Framskrivinger og 95% konfidensintervaller (i patentesen) for rådata i 2004*

	1. kvartal	2. kvartal	3. kvartal	4. kvartal
I alt	119.0 [114.0, 124.2]	122.7 [114.0, 131.9]	121.4 [110.5, 133.3]	121.9 [109.1, 136.1]
Oslo-B	119.5 [112.0, 127.4]	121.4 [108.0, 136.4]	122.1 [104.8, 142.1]	121.8 [101.7, 146.0]
Akershus	113.6 [106.8, 120.8]	117.7 [107.8, 128.5]	117.4 [105.4, 130.9]	117.6 [103.7, 133.3]
Storby	134.6 [128.4, 141.1]	139.8 [129.7, 150.7]	141.1 [128.3, 155.1]	142.6 [127.6, 159.4]
Resten	116.2 [111.3, 121.4]	120.2 [113.1, 127.8]	117.4 [109.0, 126.5]	118.0 [108.3, 128.6]

- Kvalitetsmål

Verdiene til $M1$-$M11$ er listet ut i tabell 9. Forklaringen av hvert mål er beskrevet i appendikset. Vi får en dårlig kvalitet for testen når $Mi > 1$. $M7$ har størst vekt siden dette målet beskriver sesongmønsteret i tidsserien. Når sesongvariasjoner er store eller de endrer seg over tid ($M7 > 1$), bør man ikke sesongjustere tidsserien. $M1$ og $M2$ forteller hvor store variasjoner det er i den irregulære komponenten (I_t) i forhold til den totale

Tabell 8: *Framskrivinger for sesongvariasjoner og sesongjusterte tall i 2004*

	sesongvariasjoner				se.justerte tall			
	1.kv	2.kv	3.kv	4.kv	1.kv	2.kv	3.kv	4.kv
I alt	100.3	102.1	99.3	98.5	118.7	120.2	122.3	123.8
Oslo-B	100.5	101.2	100.0	98.5	118.9	120.0	122.2	123.8
Akershus	99.3	101.8	100.1	98.9	114.4	115.7	117.4	119.0
Storby	99.9	101.5	99.7	99.0	134.8	137.8	141.5	144.1
Resten	100.1	102.5	99.0	98.4	116.1	117.3	118.7	120.0

variasjonen. Når variasjonene er store blir det vanskelig å skille trenden T_t og sesongkomponenten S_t fra I_t. For de fem tidsseriene vi analyserer i dette notatet er $M7$, $M1$ og $M2$ ganske lave. I tillegg er $M3$ nesten null. Dette kan tolkes som at endringene i I_t er små, slik at de sesongjusterte tallene ser ut som en trend.

Verdiene til $M8$-$M11$ for de tre tidsseriene "I alt", "Oslo-Bærum" og "Stavanger-Bergen-Trondheim" er store (større enn 1). Dette viser store fluktuasjoner i sesongkomponenten (S_t). Dermed kan vi få dårlige estimater for sesongjusterte serien.

Tabell 9: *Kvalitetsmål*

	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$	$M8$	$M9$	$M10$	$M11$
I alt	0.158	0.041	0.000	0.587	0.200	0.828	0.451	1.404	1.082	1.377	1.024
Oslo-B	0.321	0.048	0.000	0.587	0.200	0.842	0.675	2.168	0.976	1.524	1.047
Akershus	0.504	0.070	0.000	0.587	0.200	0.158	0.734	1.106	0.485	1.069	0.619
Storby	0.205	0.083	0.000	0.587	0.200	0.585	0.463	1.364	0.699	0.992	0.923
Resten	0.656	0.170	0.025	0.722	0.200	0.355	0.480	1.208	1.167	1.692	1.692

Tabell 10 viser gjennomsnitt og standardavvik av sesongkomponenten for de 4 kvartalene. For eksempel, for første kvartal er

$$\begin{aligned}\overline{S}_{1.kv} &= \frac{1}{12}\sum_{t=1992}^{2003} S_{1.kv,t} \\ \widehat{\sigma}^2(S_{1.kv}) &= \frac{1}{11}\sum_{t=1992}^{2003}(S_{1.kv,t} - \overline{S}_{1.kv})^2\end{aligned}$$

Vi ser at standardavvikene er veldig lave i forhold til gjennomsnittene. Dette indikerer at sesongvariasjonene er stabile. Vi får dermed lave revisjoner for sesongjusterte tall.

Vi har en annen diagnotisk prosedyre for å vurdere kvaliteten til en sesongjustering. Det er "sliding spans". Metoden er utviklet av Findley (1990) et al. og har vært brukt i X12-ARIMA. Findley og Monsell har gitt eksempler som viser at sliding spans gir bedre evaluering av kvaliteten enn de 11 målene $M1$-$M11$.

- Sliding spans

Formålet er å se de største variasjonene av sesongkomponenten og sesongjusterte tall når data forlenges. Avhengig av lengden til en tidsserie, lager X12-ARIMA fra de opprinnelige

Tabell 10: *Gjennomsnitt og standardavvik (i parentes) for 4 kvartaler*

	1.kvartal	2.kvartal	3.kvartal	4.kvartal
I alt	99.08 (0.63)	101.83 (0.83)	100.25 (0.71)	98.71 (0.67)
Oslo-B	99.70 (1.15)	101.74 (1.07)	100.29 (0.74)	98.10 (1.27)
Akershus	98.81 (0.22)	101.75 (0.76)	100.35 (0.31)	99.00 (0.40)
Storby	99.52 (0.34)	101.49 (0.71)	100.47 (0.74)	98.45 (0.36)
Resten	98.60 (0.65)	101.85 (0.78)	100.36 (0.87)	99.02 (0.52)

data opptil 4 overlappende delmengder (såkalt "span" på engelsk). Hver delmengde er betraktet som en komplett tidsserie, dvs effektene av ukedager, påske og intervensjoner blir korrigert før man sesongjusterer dem. De 4 følgende delmengdene blir valgt.

[1kv.94 - 4kv.2000], [1kv.95 - 4kv.2001], [1kv.96 - 4kv.2002], og [1kv.97 - 4kv.2003]

La $S_t(k)$ og $A_t(k)$ være sesongfaktor og sesongjusterte tall som er estimert i tidspunkt t fra delmengden k. $E_t(k)$ er endring i prosent fra $t-1$ til t av sesongjusterte tall i delmengden k, $E_t(k) = (A_t(k) - A_{t-1}(k))/A_{t-1}(k)$.

Vi sier at sesongkomponenten i tidspunkt t er upålitelig hvis den største relative endringen $S_t^{max} = (max_k S_t(k) - min_k S_t(k))/min_k S_t(k)$, er større enn 0,03. Vi definerer på samme måte for sesongjusterte tall.

Vi sier at endringen i prosent fra tidspunkt $t-1$ til t, i den sesongjusterte serien, er upålitelig hvis den største differansen i prosent $E_t^{max} = max_k E_t(k) - min_k E_t(k)$, er større enn 0,03,

Grenseverdien for prosentandelen tidspunkter hvor sesongkomponenten er definert som upålitelig, er 15%, og for prosentandelen tidspunkter hvor $E_t^{max} > 0.03$, er 35-40%.

Resultatene viser at det er ingen tidspunkt der S_t^{max} eller E_t^{max} er større enn 0.03. Vi får gode sesongjusteringer for alle 5 tidsseriene.

- Figurer

Skala for sesongkomponenten. Vi sesongjusterer de 5 tidsseriene med multiplikativ model

$$Y_t = T_t \times S_t \times I_t \tag{2}$$

I modellen (2) har Y_t og T_t samme enhet, men S_t og I_t er beregnet i prosent. Siden Y_t er indeksen, har vi ikke en enhet for data og trenden. Vi kan dermed plotte S_t og I_t med den skalaen som vi har brukt for rådata eller vi lar S_t og I_t varierer i intervallet $[min(S_t, I_t),\ max(S_t, I_t)]$ på y-aksen. De to metodene kan gi to forskjellige bilder av sesongkomponenten.

- Vi plotter sesongkomponenten av tidsserien "I alt" i figur 4 med den skalaen i figur 1, fra 30 til 140. På grunn av at sesongvariasjonene er ganske lave (se tabell 9), er sesongmønsteret helt flat. Vi kan ikke se endringene fra tidspunkt til tidspunkt eller fra år til år i figur 4. Det gir bare et bilde av S_t i forhold til Y_t.

- Vi vil velge en smalere skala for sesongkomponenten, fra 96 til 104. Det plotter vi i figur 5. Vi ser et klart sesongmønsteret. Skalaen fra 96 til 104 skal brukes for sesongkomponenten for alle tidsseriene.

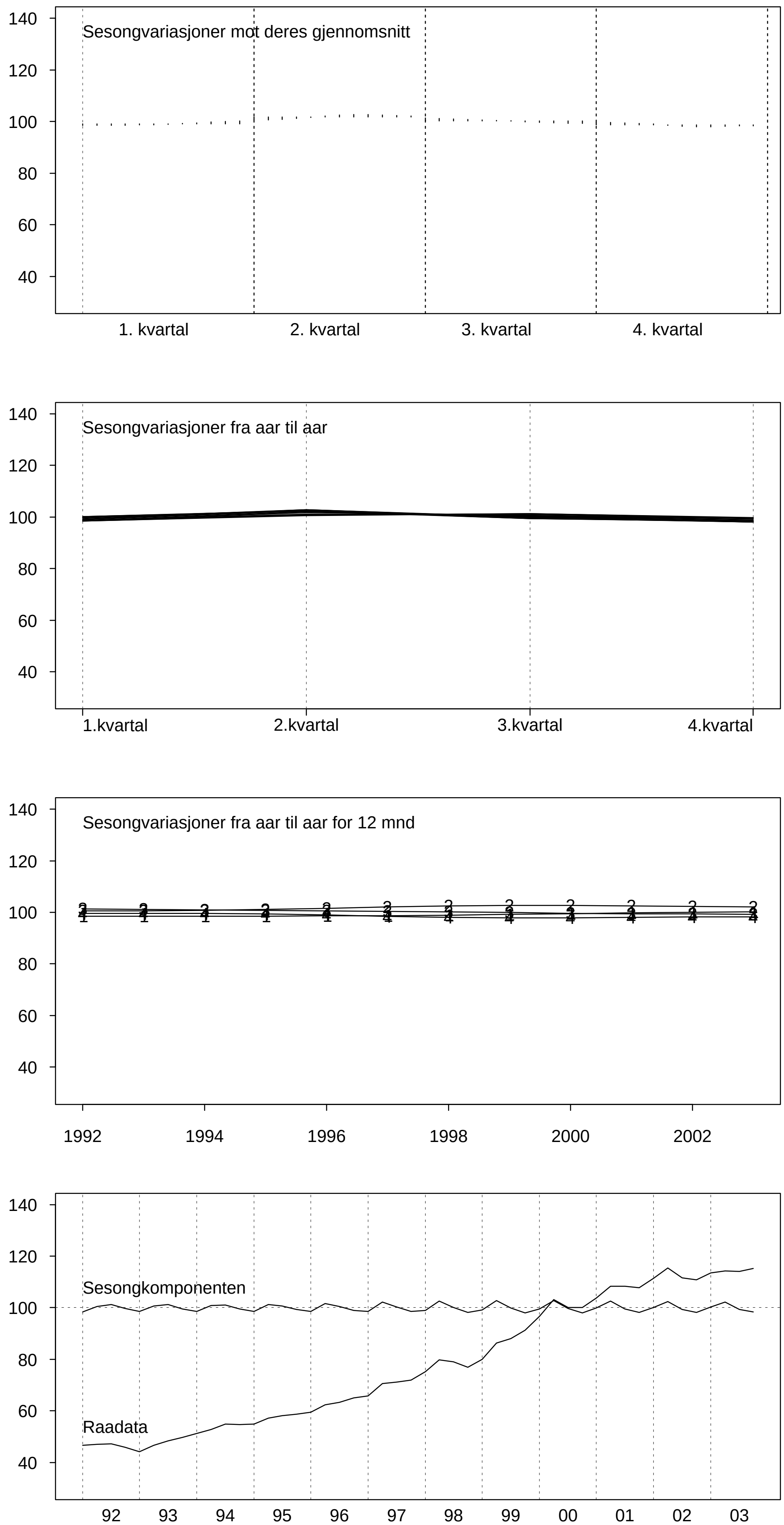
Figur 5-9 er sesongkomponentene for 5 tidsserier. Det første plottet i figuret er sesongvariasjoner mot deres gjennomsnitter for 4 kvartaler. Det andre er sesongmønstre fra år til år. Det tredje er sesongvariasjoner for 4 kvartaler gjennom årene. Det fjerde er det vanlige plottet.

De 5 tidsseriene har akkurat samme sesongmønster. I gjennomsnitt er andre kvartalet høyeste (se første plott i figuren). Det er en vri i sesongkomponenten mellom andre og tredje kvartalet (se andre plot). Dette skyldes at i perioden 1992-1994 var aktiviteten i boligmarkedet størst i tredje kvartal, men i perioden 1994-2003 er aktiviteten i andre kvartal størst. Det kom etter i rekkefølge er andre, fjerde og første kvartal. Forskjellen mellom deres nivåer er ikke så stor. Etter 1995 er det en kraftig økning i aktiviteten i andre kvartal som skyver dette kvartalet ut fra de andre. I hele året er det fjerde kvartal som har minst aktiviteten, særlig i Oslo og Bærum. (se tredje og fjerde plott).

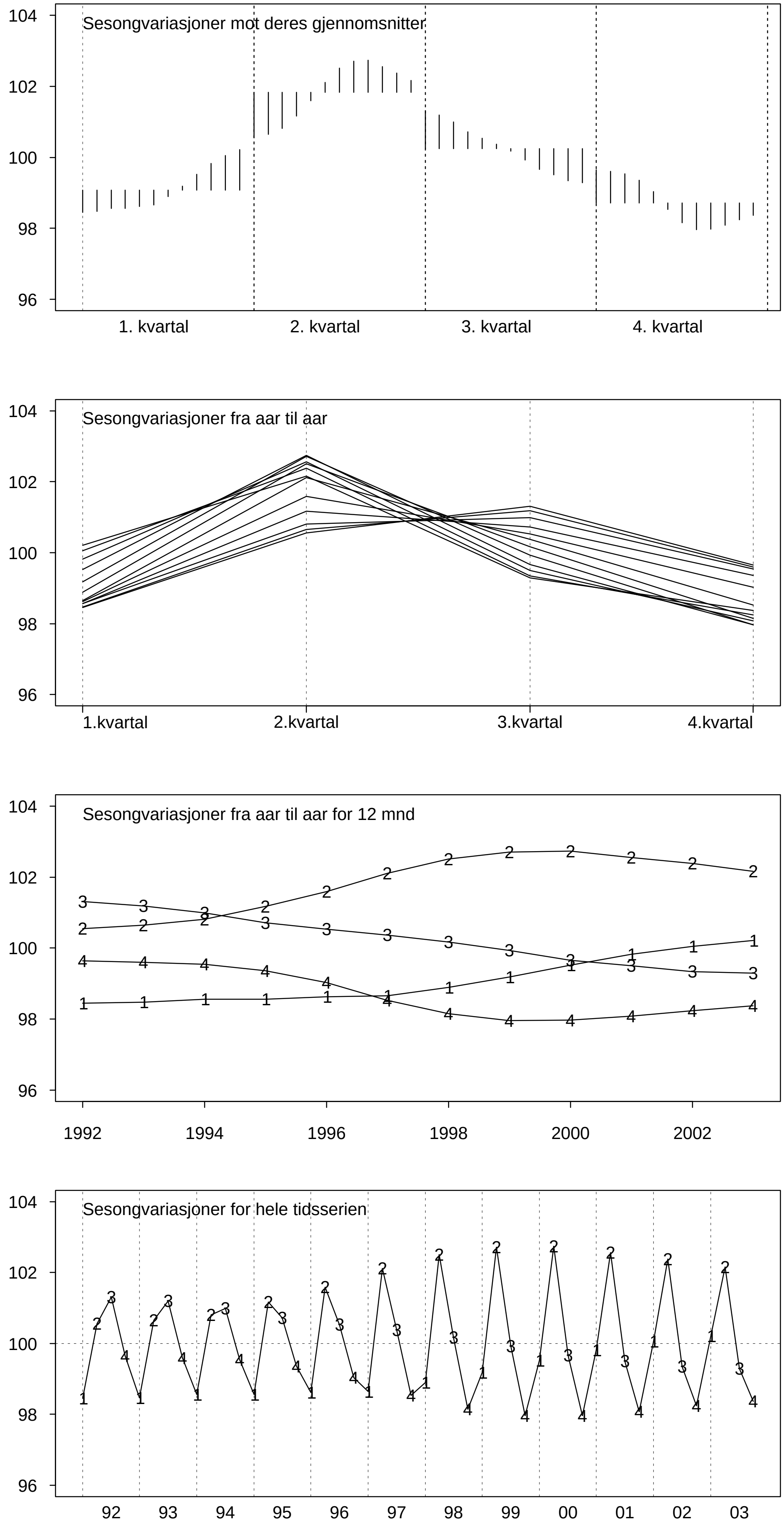
Figur 10-14 viser trend og sesongjusterte tall for 5 tidsserier. Siden variasjoner av den irregulære komponenten er veldig små blir trenden og de sesongjusterte tallene nesten identiske. Sesongvariasjonene er også lave, dermed er rådata og de sesongjusterte tallene er omtrent like.

4 Oppsummering

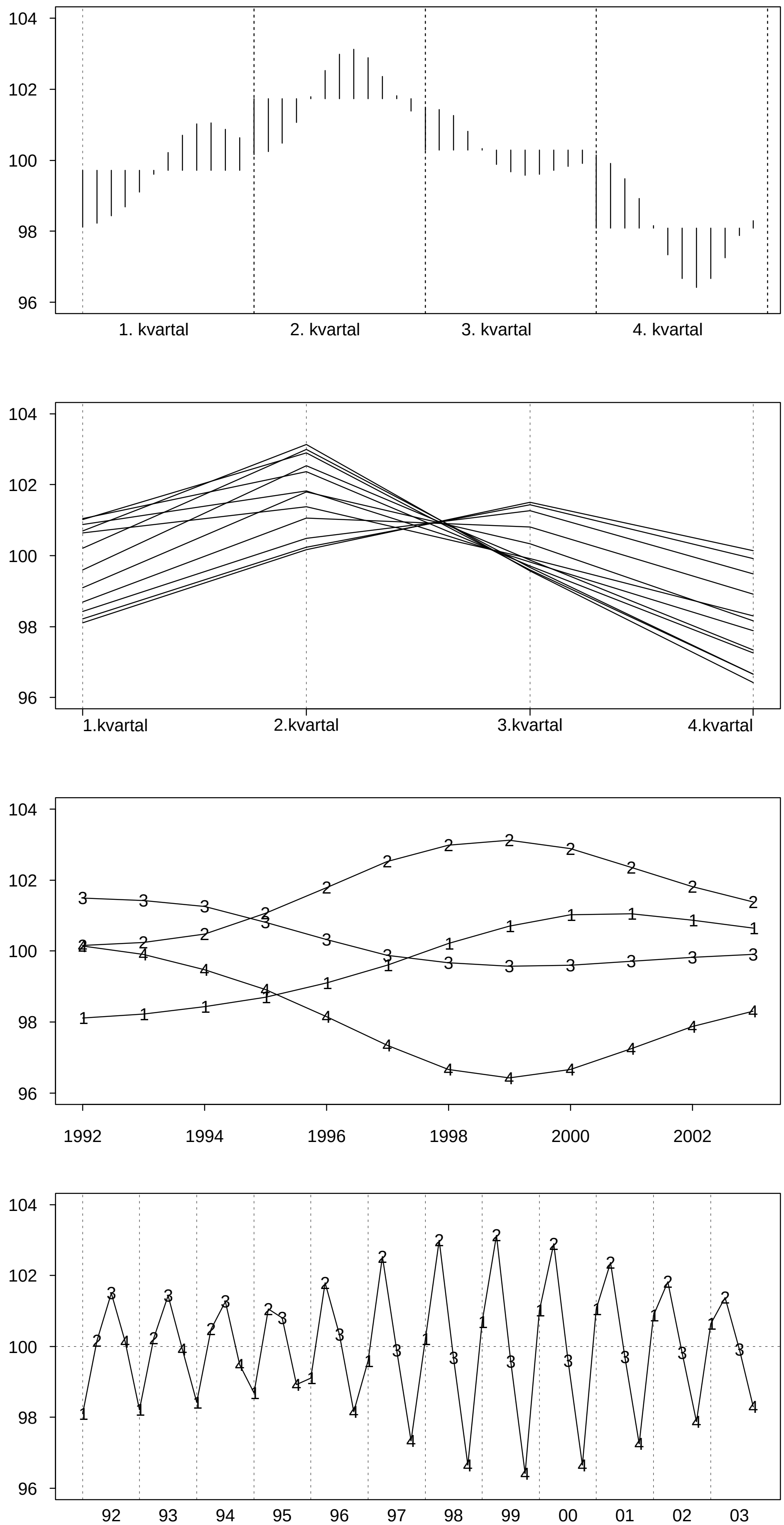
Vi har sesongjustert boligprisindeksen for 5 tidsserier "I alt", "Oslo-Bærum", "Akershus", "Storby" og "Resten". Vi ser at prisene har økt over hele landet og mest i Stavanger, Bergen og Trondheim de siste årene. Testen viser at det er sesongvariasjoner i rådata med topp i andre kvartal. Kvalitetsmål og 'sliding spans' indikerer en god sesongjustering i den forstand at revisjoner av de sesongjusterte tallene er små når nye observasjoner legges til. Variasjonene i den irregulære komponenten er veldig lave slik at de sesongjusterte tallene ser ut som en trend. Vi kan konkludere med at de fem tidsseriene er velegnet for sesongjustering.



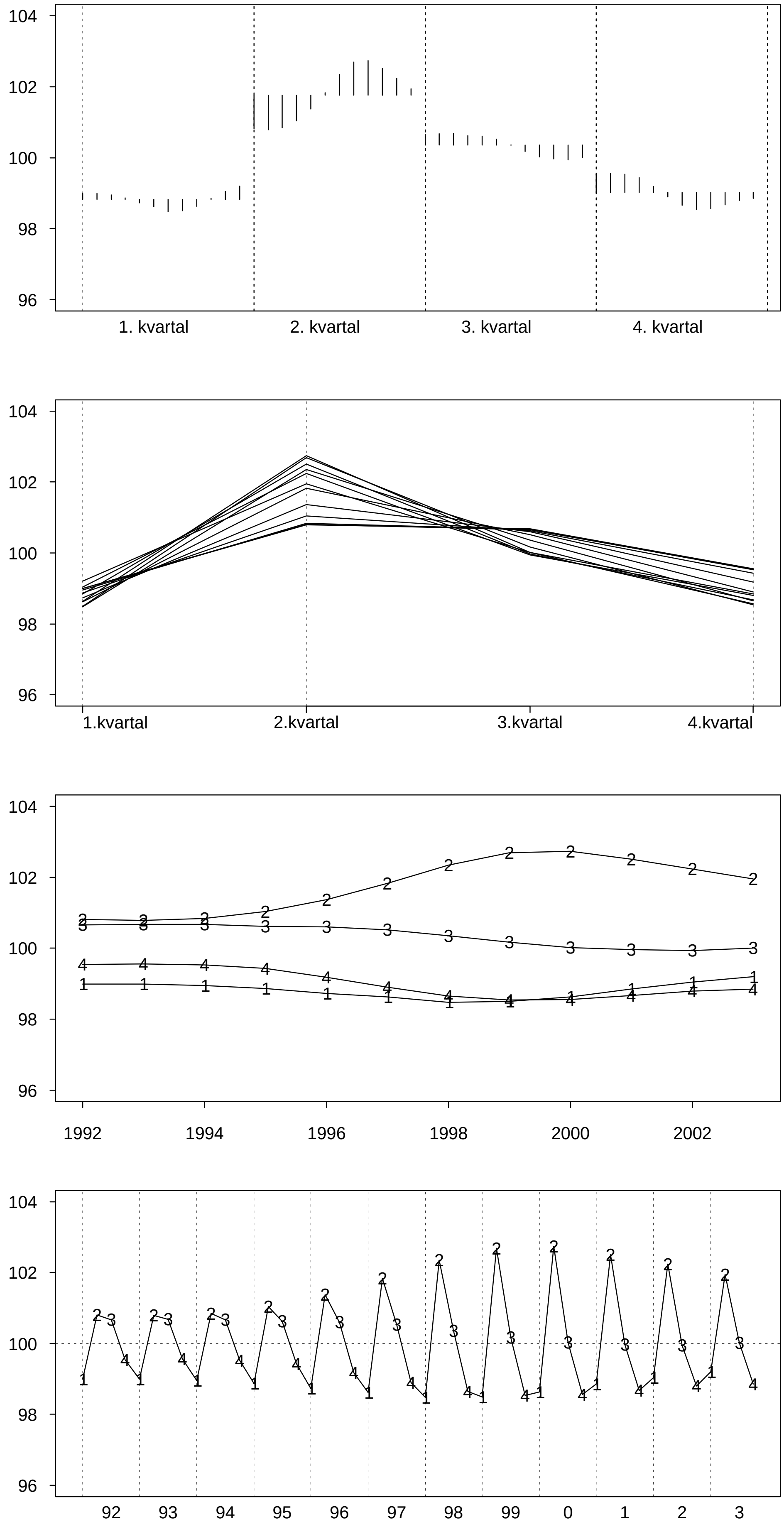
Figur 4: *I alt. Sesongkomponent med en skala fra 30 til 140*



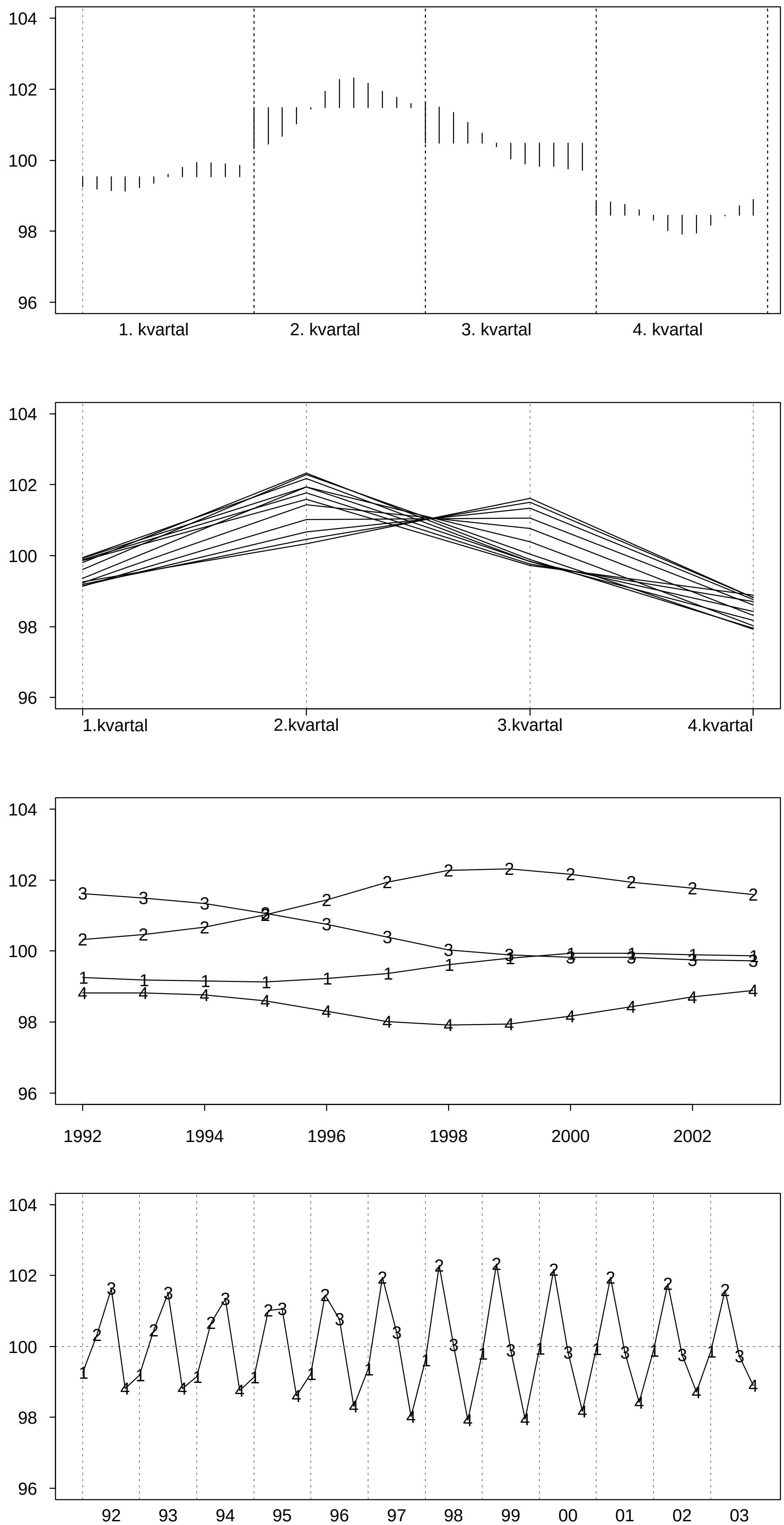
Figur 5: *I alt. Sesongkomponent med en skala fra 96 til 104*



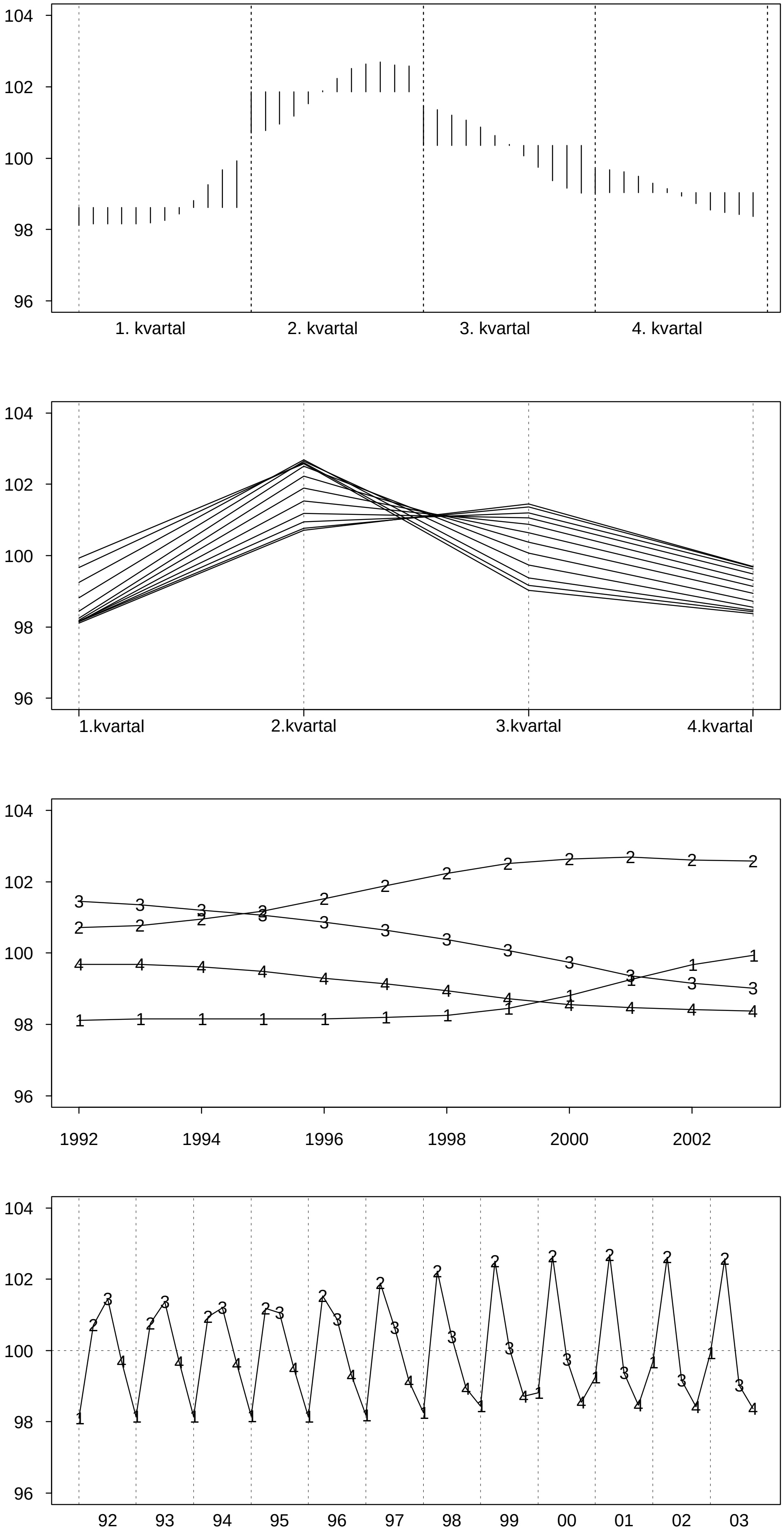
Figur 6: *Oslo og Bærum. Sesongkomponent*



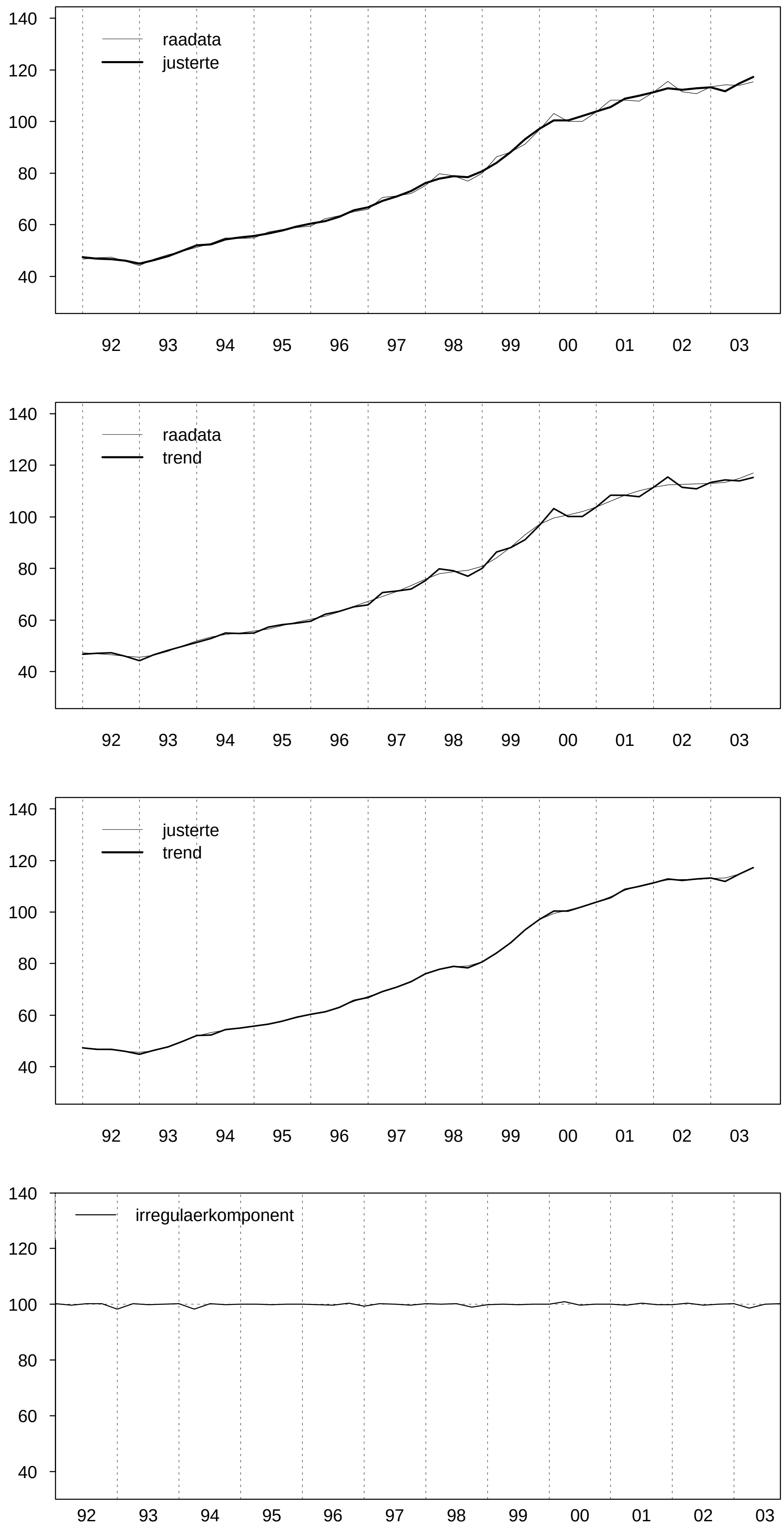
Figur 7: *Akershus utenom Bærum. Sesongkomponent*



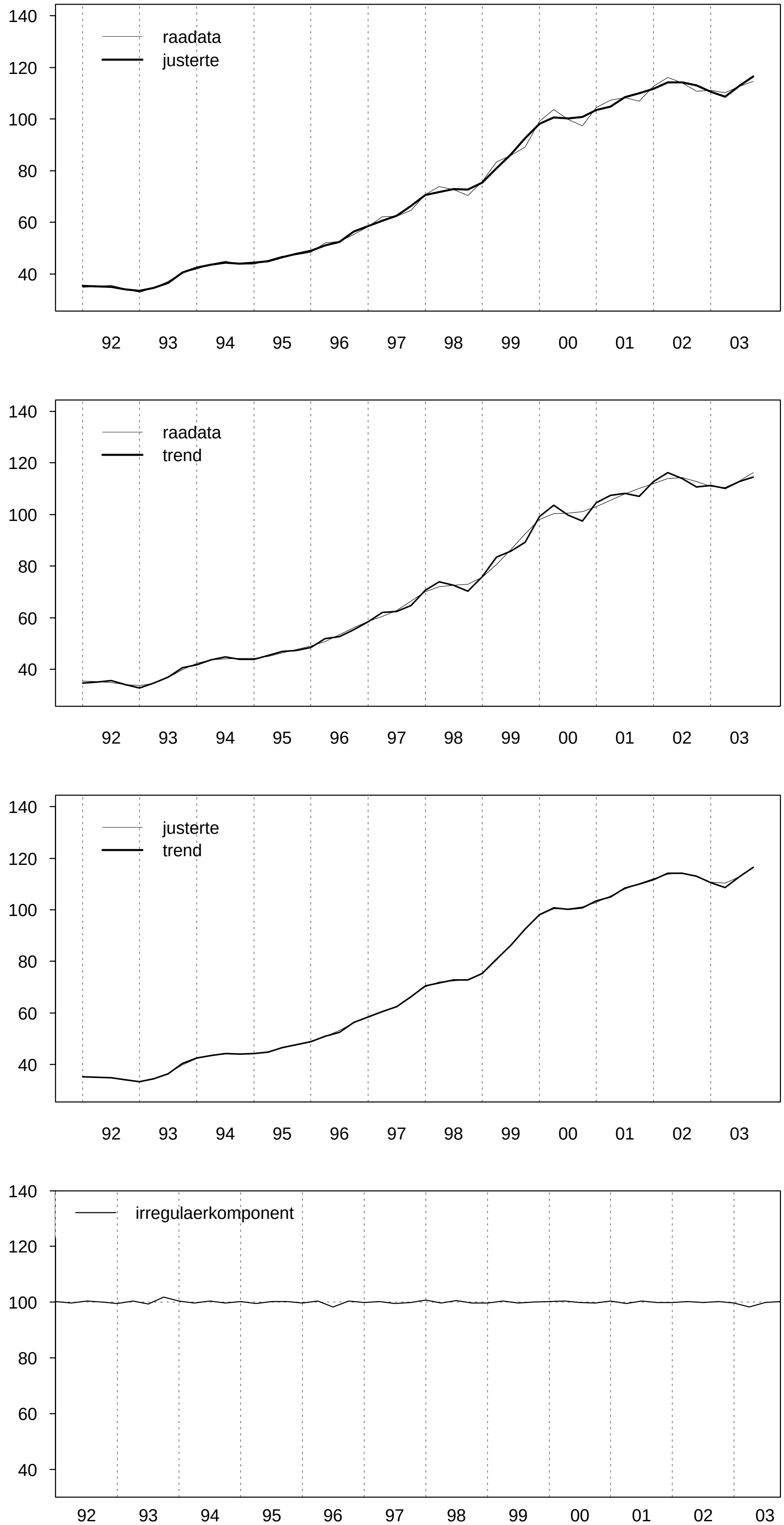
Figur 8: *Stavanger-Bergen-Trondheim. Sesongkomponent*



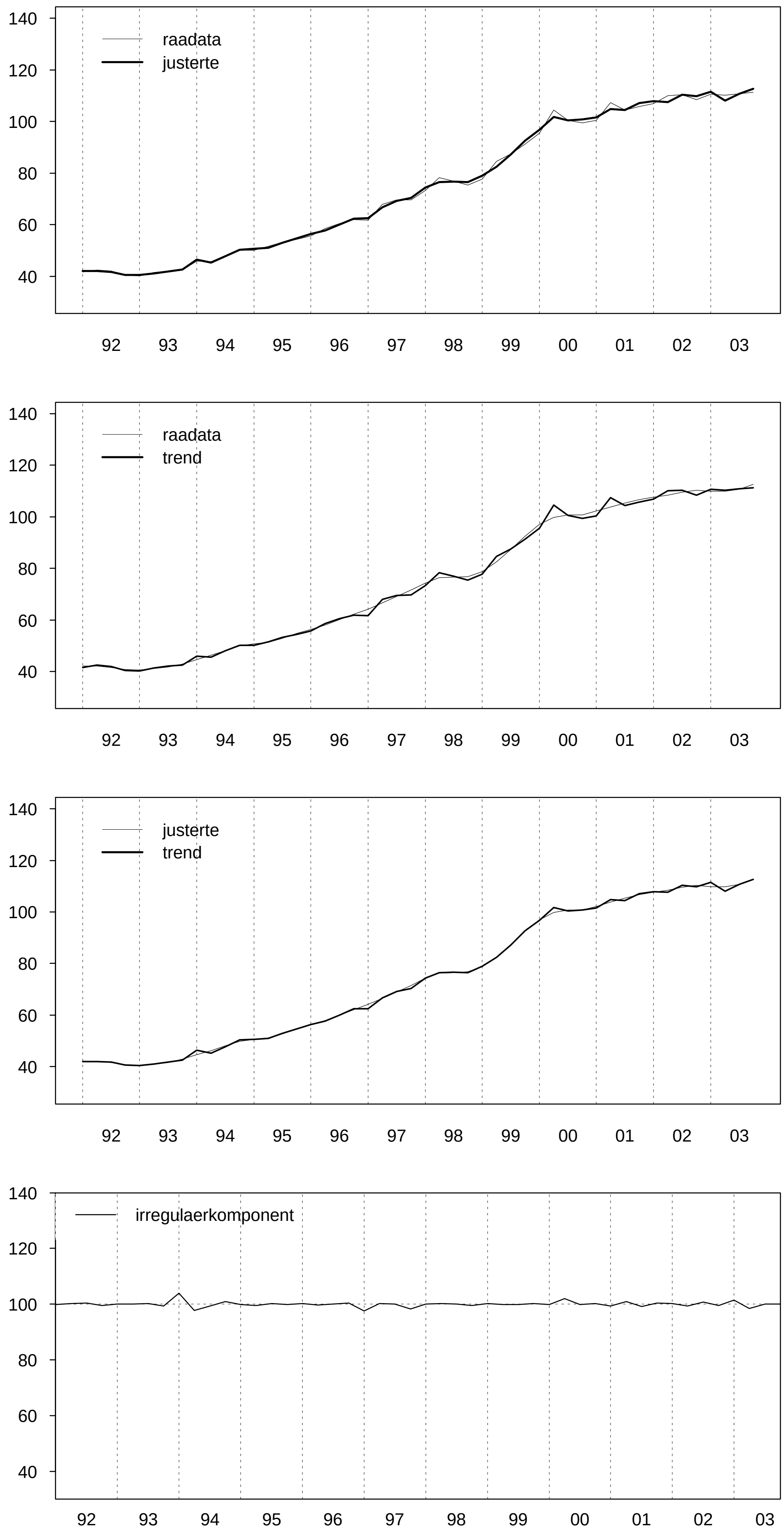
Figur 9: *Resten av landet. Sesongkomponent*



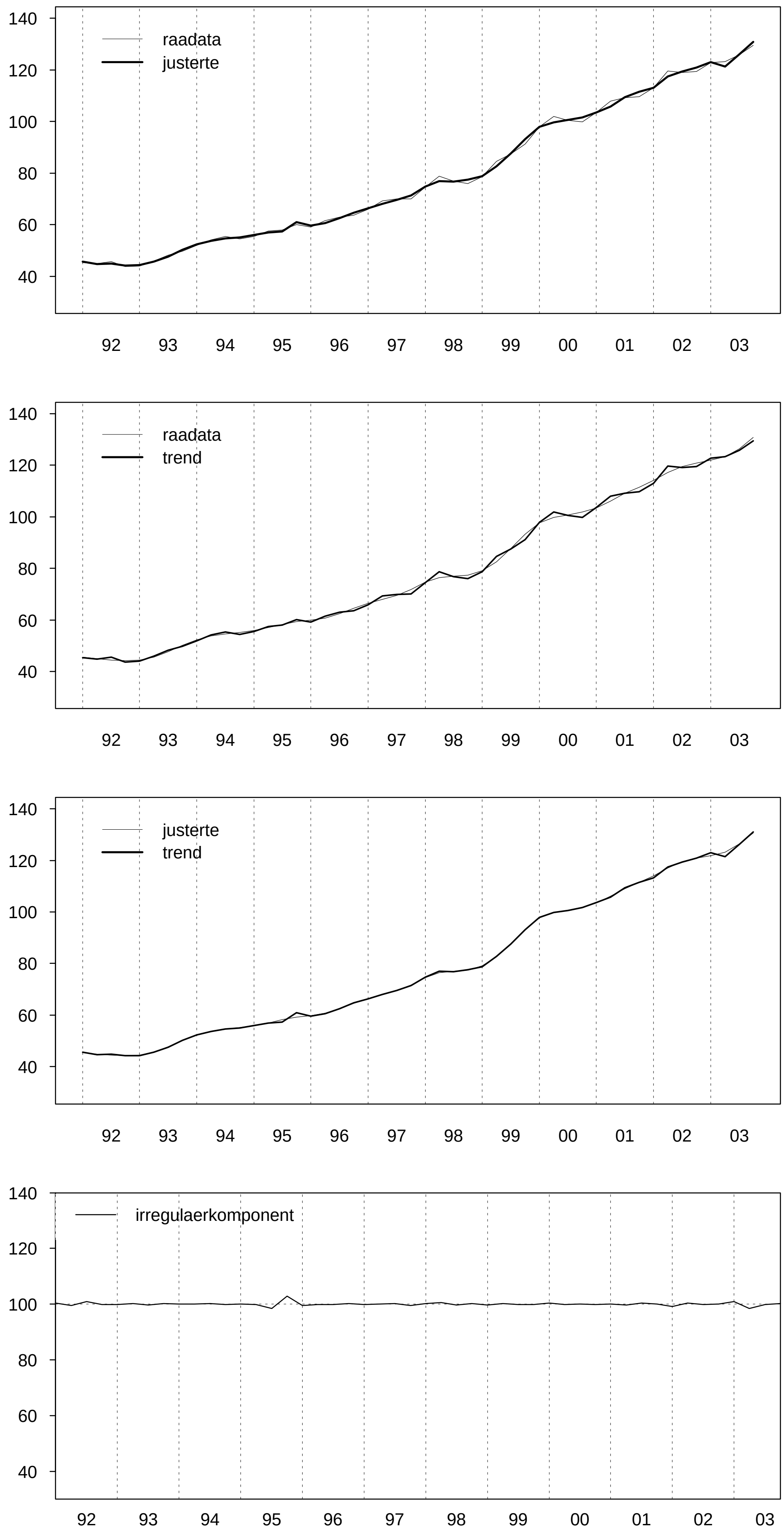
Figur 10: *I alt. Trend og sesongjusterte tall*



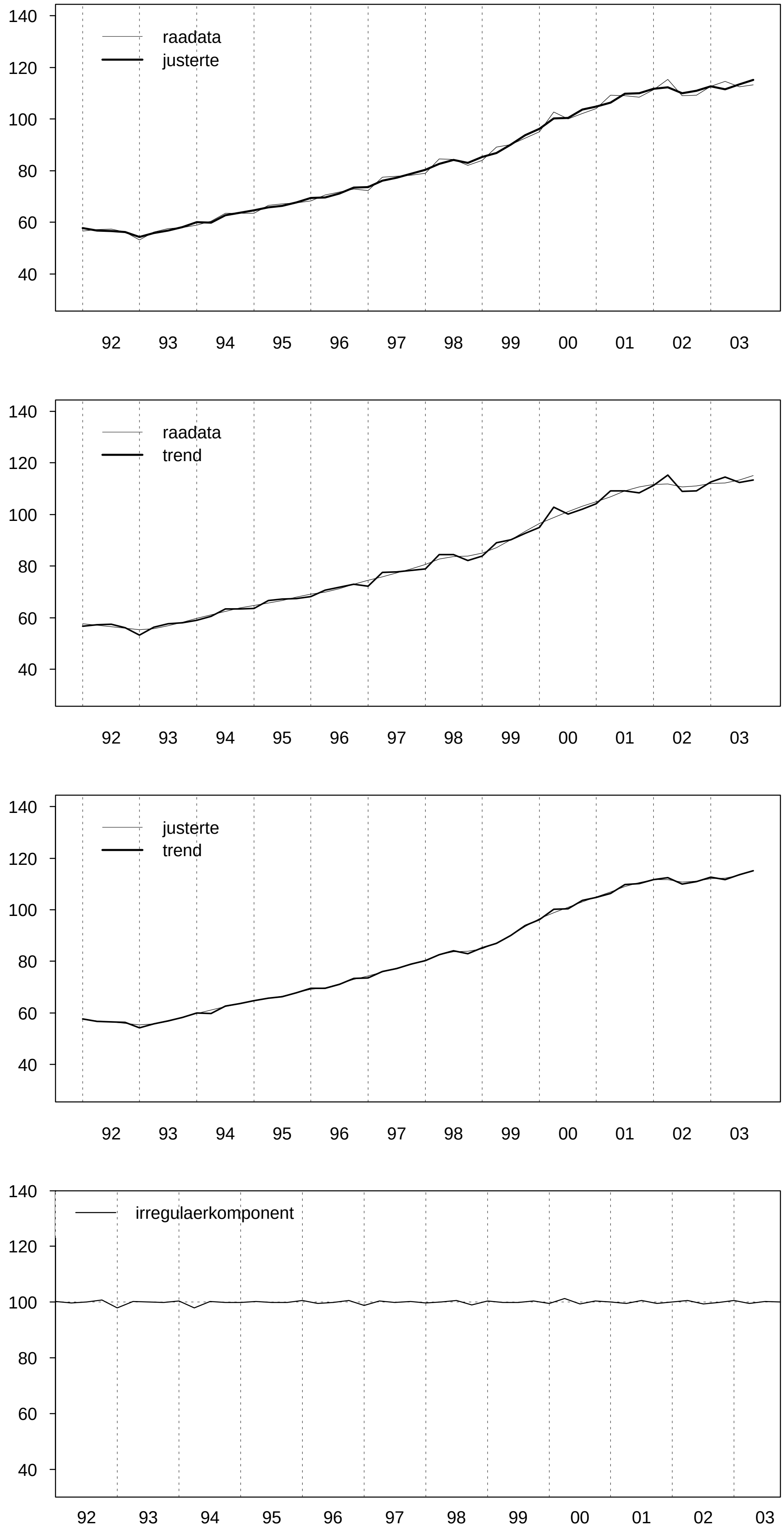
Figur 11: *Oslo og Bærum. Trend og sesongjusterte tall*



Figur 12: *Akershus utenom Bærum. Trend og sesongjusterte tall*



Figur 13: *Stavanger-Bergen-Trondheim. Trend og sesongjusterte tall*



Figur 14: *Resten av landet. Trend og sesongjusterte tall*

Referanser

[1] Alan Pankratz (1991), Forecasting with Dynamic Regression Models", Wiley Interscience

[2] Bell W. R. and Hillmer S. C. (1983). "Modelling Time Series With Calendar Variation". *Journal of the American Statistical Association*, 78, 526-534

[3] Cleveland W. S. and Susan J. D. (1980), "Calendar Effects in Monthly Times Series: Detection by Spectrum Analysis and Graphical Methods", *Journal of the American Statistical Association*, 75, 487-495

[4] Findley D. F., Brian C. Monsell, William R. Bell, Mark C. Otto and Bor-Chung Chen (1998). New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program", Journal of Business & Economic Statistics, 16, 127-177

[5] Dagum Estela Bee, Benoir Quenneville and Brajendra Sutradhar (1992). Trading-day Variations Multiple Regression Models with Random Parameters", International Statistical Review, 60, 57-73

[6] Dagum Estela Bee (1988). The X11ARIMA/88 Seasonal Adjustment Method Foundations and User's Manual"

[7] John Higginson (1975). "An F Test for the presence of moving seasonality when using census method II-X-11 variant"

[8] Lars A. Loe (1987). Framskriving av tidsseriedata i kvartalsvis nasjonalregnskap", Notater 87/1

[9] Leiv Solheim og Dinh Quang Pham (1997). Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997", Notater 73/97

[10] Lothian J. and M. Morry. "A set of Quality Control Statistics for the X-11 ARIMA"

[11] Bureau of the Census. "X-12 ARIMA Reference Manual, Version 0.2.5, October 1, 1999"

[12] S-PLUS User'Manual, version 3.2, December 1993

Appendiks

Kvalitets mål i X-11 ARIMA

X-11 ARIMA programmet lister ut i tabell F.2 en liste med 11 mål $M1 - M11$, hvor deres verdier varierer fra 0 til 3. For verdier under 1 er kvaliteten tilfredstillende. Kvaliteten blir dårligere når M_i går mot 3. Formlene for $M1 - M11$ er gitt i notatet av J. Lothian og M. Morry (1978.c). La T_t, S_t og I_t være trenden, sesongkomponenten og irregulærkomponenten i en tidsserie. A_t er den sesongjusterte serien. De 11 målene beskrives i det følgende.

- $M1$: *Relativt bidrag fra I_t i den totale variasjonen.* La

  $$\overline{I}(k)^2 = \frac{1}{N-k} \sum_{t=k+1}^{N} (I_t - I_{t-k})^2$$

  $\overline{I}(k)^2$ er den totale variasjonen i lag k fra I_t. For de originale dataene er beregnet ved

  $$\overline{O}'(k)^2 = \overline{T}(k)^2 + \overline{S}(k)^2 + \overline{I}(k)^2$$

  der $\overline{T}(k)^2$ og $\overline{S}(k)^2$ beregnes på samme måte som $\overline{I}(k)^2$, fra tabellene D12 og D10. Det relative bidraget fra I_t i forhold til originale dataene er

  $$R_{I(k)} = \frac{\overline{I}(k)^2}{\overline{O}'(k)^2} \times 100\%$$

  Ved erfaringer viser man at $k = 3$ er best for å sammenligne. Grenseverdien for $R_{I(k)}$ er 10%. La

  $$M1 = \frac{R_{I(3)}}{10}$$

  så grenseverdien for $M1$ er 1. Når $M1 > 1$ vil det vise at $\overline{I}(k)^2$ er høy i forhold til $\overline{O}'(k)^2$, slik at trend eller sesongkomponent kan bli forstyrret av variasjoner fra I_t. Dermed kan hverken T_t eller S_t skilles helt perfekt fra I_t og sesongjusteringen er av lavere kvalitet.

- $M2$: *Relativt bidrag fra I_t til den stasjonære delen.* Bidraget av I_t sammenlignes med den originale serien etter å ha fjernet trendkomponenten fra den første runden. En størrelse $M2$ er innført for denne testen. Hvis $M2 > 1$ vil det vise at bidraget fra I_t er høyt.

- $M3$: *Endring fra måned til måned i I_t i forhold til endring i T_t.* Formålet i sesongdekomponeringen er å fjerne sesongkomponenten fra den originale serien for å ha et godt bilde av sesongjusterte serien. Ikke bare det, trenden og den irregulære komponenten må også være helt klar. Hvis endringen fra måned til måned i I_t er større enn endringen i trenden, blir det vanskelig å skille de to komponentene I_t og T_t i den tredje runden. Kvaliteten til dekomponeringen blir mindre. En størrelse $M3$ er definert for denne testen. Hvis $M3 > 1$ er variasjonen til I_t for høy.

- $M4$: *Mål på autokorrelasjon i I_t.* I sesongjustering ønsker man at I_t er en uavhengig og tilfeldig prosess. Man antar at I_t skrives på formen $I_t = \rho I_{t-1} + \epsilon_t$, ($\epsilon_t$ er hvit støy prosess). Testen er $H_0 : \rho = 0$ mot $H_1 : \rho \neq 0$. Et mål $M4$ er innført for denne testen. Når $M4$ er større enn 1, vil det vise en signifikans for $\rho \neq 0$. Da er I_t ikke en uavhengig tilfeldig prosess. Dette medfører at alle tester som er basert på I_t kan være ugyldige.

- $M5$: *Antallet måneder for at endringen i T_t skal bli større enn endringen i I_t.* En vil teste at det fins en k slik at etter k måneder skal endringene i T_t dominere endringene i I_t. Testen er ikke signifikant når $M5 > 1$.

- $M6$: *Endring fra år til år i I_t i forhold til endringen i S_t.* For å skille de to komponentene I_t og S_t bruker man (3×5) filtret for SI komponenten i den andre runden. Ved erfaringer viser det at når forholdet $\overline{I}/\overline{S}$ er for lavt $(< 1,5)$, er (3×5) filtret ikke tilpasset, siden filtret ikke er fleksibelt nok til å følge sesongvariasjonen. Omvendt når $\overline{I}/\overline{S}$ er for høyt $(> 6,5)$, da er det (3×5) filtret for fleksibelt slik at sesongkomponenten blander seg i den irregulære komponenten. En størrelse $M6$ er innført for å gi opplysninger til filteret. Hvis $M6 > 1$ betyr det at (3×5) filtret ikke er tilpasset. Et alternativ er å bruke (3×1) filtret når forholdet $\overline{I}/\overline{S} < 1,5$ eller stabil opsjonen når $\overline{I}/\overline{S} > 6,5$. Forholdet $\overline{I}/\overline{S}$ er listet ut i tabell F2.H.

- $M7$: *Grad av bevegelig sesongmønster i forhold til stabilt sesongmønster.* Formålet er å måle hvor stor andelen av den stabile delen er i forhold til bevegelige delen i sesongkomponenten. En lager et mål $M7$ ved å basere seg på to F-tester, en for stabiliteten F_S og en for bevegelsen F_M. Lave verdier i F_S og høye verdier i F_M vil medføre til at sesongmønsteret ikke er identifiserbart og verdien til $M7$ øker. Grensen for $M7$ er 1. Når $M7 > 1$ sier vi at sesongmønsteret ikke er identifiserbart.
  Blant 11 tester for kvaliteten er denne testen viktigst. Årsakene til at $M7 > 1$ er:

  (i) Ingen sesong i det hele tatt i den originale serien.

  (ii) Variasjonen i sesongkomponenten er så stor at sesongmønsteret ikke kan identifiseres.

  (iii) Additiv modell er brukt når serien tilpasses bedre med multiplikativ modell.

  (iii) kan rettes ved å endre modell for den originale serien. Verdien til $M7$ kan da bli mindre.

  Ofte vil man ikke sesongjustere en serie når verdien til $M7$ er større enn 1.

- $M8$: *Grad av fluktuasjoner i sesongkomponenten gjennom hele serien.* En størrelse $M8$ er innført for måle fluktuasjoner i den sesongkomponenten S_t. $M8 > 1$ viser at fluktuasjonene er så store at S_t ikke er stabil lenger. Dermed får en et dårlig estimat for den sesongjusterte serien.

- $M9$: *Grad av lineær bevegelse i S_t i hele serien.* $M9 > 1$ viser at fluktuasjoner i S_t ikke er tilfeldige.

- $M10$: *Grad av fluktuasjoner i S_t i de siste årene.* Denne testen er samme som $M8$ men bare for de siste årene.

- $M11$: *Grad av lineær bevegelse i S_t i de siste årene.* Denne testen er samme som $M9$ men bare for de siste årene.

De to målene $M10$ og $M11$ gir informasjon om fluktuasjoner og bevegelser i sesongkomponenten for de siste årene. Testen for grad av fluktuasjoner ($M8$ og $M10$) er mindre viktig enn testen for grad av bevegelser ($M9$ og $M11$) i sesongkomponenten. Til slutt har vi en størrelse Q som er definert slik

$$Q = \frac{\sum_{i=1}^{11} w_i Mi}{\sum_{i=1}^{11} w_i}$$

hvor w_i er vekten for M_i. Målene som er relativt viktige, har større vekt. De 11 w_i-ene er gitt i tabell 1.

Tabell 1. 11 vekter for 11 mål.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
Verdi	13	13	10	5	11	10	16	7	7	4	4

Vi ser at $M1$, $M2$ og $M7$ har største vekter fordi de er relativt sett viktigere enn de andre. Det er noen punkter som vi synes er viktige for brukeren for å kunne vurdere kvaliteten til den sesongjusterte serien.

- De 11 målene gjelder for de fleste seriene, men ikke for alle. De er bare empiriske mål.
- Mål som har større vekt er relativt sett viktigere.
- Hvis en serie der alle $Mi > 1$ vil vi ikke sesongjustere denne serien.
- Ingen mål kan alene bestemme kvaliteten av sesongdekomponeringen.
- Når $M7 > 1$ skal en være svært forsiktig med å sesongjustere serien.
- Hvis brukeren bruker ett annet filter enn (3×5) filtret for estimering av sesongfaktorene, er $w_6 = 0$.
- For en serie som er mindre enn 6 år eller stabil opsjon er brukt for estimering av sesongfaktorer, får vektene w_8, w_9, w_{10} og w_{11} null verdiene. De 11 modifiserte vektene for de 11 målene er vist i tabell 2.

Tabell 2. 11 modifiserte vektene.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
Verdi	17	17	10	5	11	10	30	0	0	0	0

Programmer

1. I alt

Dette er programmet for å sesongjustere tidsserien "Ialt" i X12-ARIMA. Input-filen er en flat fil med 5 kolonner som står for i alt, Oslo og Bærum, Akerhus utenom Bærum, Stavanger-Bergen-Trondheim, og resten av landet, med format (F7.1,F7.1,F7.1,F7.1,F7.1). Vi korrigerer ikke rådata for effektene av ukedager og påske. ARIMA modellen velges selv av programmet. Ekstremverdier og framskrivinger beregnes med den modellen som blir valgt. "Sliding spans" brukes for å beregne revisjoner av sesongjusterte tall. Utfilene er trend, sesongkomponenten, sesongjusterte tall og den irreglære komponenten.

```
series{title="Boligprisindeksen i alt"
       start=1992.1              # dato til den første observasjonen
       period=4                  # det er kvartalstall
       file="bolig1.dat"         # input-fil
       format="(F7.1,28X)"       # format av input-filen
       decimals=1}               # antall desimaler
transform{function=log}          # sesongjustering med multiplikativ modell
regression{}                     # regresjonsanalyse
automdl{file="my.mdl"            # programmet finner selv en ARIMA modell
identify=all
savelog=(automodel)}
forecast{save=(fct)}             # framskrivinger
outlier{}                        # ekstremverdier
estimate{maxiter=100}
x11{savelog=(m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 q)
    print=all
    save=(d10 d11 d12 d13)}
slidingspans{}
```

Programmet som benyttes for å sesongjustere de andre tidsseriene ser tilsvarende ut.

De sist utgitte publikasjonene i serien Notater

2003/92 J.I. Hamre: Undersøkelsen om legemeldt sykefravær. Dokumentasjon av utvalgsplan, trekking og rullering for 2003. 37s.

2004/1 A.G. Pedersen: Sammenligning av manuell og auomatisert metode ved koding av dødsårsak. 22s.

2004/2 T.M. Köber: Registerbasert sysselsettingsstatistikk for helse og sosialhjelp. 42s.

2004/3 T. Dypbukt: Tilpasningseffekter av utbytteskatten i 2000/2001. 38s.

2004/4 A.H. Foss: Kvaliteten i arbeidsmarkedsdelen i Folke- og boligtellingen 2001. 42s.

2004/5 L.C. Zhang: Domene-estimering i lønnsstatistikk. 14s.

2004/6 J. Kjelvik: Del I: Kommunenes utgifter til primærlegetjenesten 2002. Del II: Organisering av legevakttjenesten. 52s.

2004/7 K. Olsen: Forsystem for ikke-finansielle foretak i nasjonalregnskapet, dokumentasjon av teknisk drift. 29s.

2004/8 K. Olsen: Database for de institusjonelle sektorene i nasjonalregnskapet, dokumentasjon av teknisk drift. 24s.

2004/9 K. Olsen: Forsystem for finansielle foretak i nasjonalregnskapet, dokumentasjon av teknisk drift. 30s.

2004/10 T. Bye, P.R Johansen og K.G Salvanes: Evaluering av Arbeidstilbudsforskningen i SSBs forskningsavdeling. 119s.

2004/11 A. M. Auno, B. Gabrielsen, T. Hagen, T. Kvalø og K. Vetvik: ILO-Arbeids-kraftregnskap. Delprosjekt arbeidstid. 44 s.

2004/12 K. Lorenzen: Dokumentasjon av registrering av selvstendige i 2000 - kriterier for opplasting og oppfølging etter opplasting. 41s.

2004/13 S. Flåte, B.O. Lagerstrøm og E. Wedde: Barns levekår i lavinntektsfamilier. Dokumentasjonsrapport. 68s.

2004/14 D.Q. Pham: Korrigering for helligdager for ukeverk i AKU. 27s.

2004/15 T.M. Normann: Omnibusundersøkelsen november/ desember 2003. Dokumentasjonsrapport. 49s.

2004/16 A. Sundvoll og L. Taule: Utviklingsprosjekt for kirkelig tjenestestatistikk. Dokumentasjonsrapport. 51s.

2004/17 S. Flåte: Undersøkelse om trygghet i hverdagen. Dokumentasjonsrapport. 46s

2004/18 H.C.Hougen og C. Wiecek: Undersøkelse om levekår og psykisk helse blant innsatte i norske fengsler. Dokumentasjonsrapport. 94s.

2004/19 E.Eng Eibak: Forventningsindikator - konsumprisene. November 2003-mai 2004. 23s.

2004/20 V.V. Holst Bloch, E. Engelien og M. Steinnes. Arealklassifikasjon i tettsteder. En uttesting av nasjonal arealdekkeklassifikasjon i deler av Fredrikstad tettsted. 55s.

2004/21 A. Holmøy og E.Wedde: Undersøkelse om arbeid, livsstil og helse 2003. Dokumentasjonsrapport. 38s.

2004/22 H.C. Hougen og M.A. Gløboden: Samordnet levekårsundersøkelse 2002-tverrsnittundersøkelsen. Dokumentasjonsrapport. 110s.

2004/23 H. Utne: Håndbok for Folke- og boligtelling 2001. 63s.

2004/24 A. Holmøy: Undersøkelse om livsløp, aldring og generasjon (LAG). Dokumentasjonsrapport. Oppdatert versjon av Notat. 2003/88. 129s.