

# En studie av sammenhengen mellom makroøkonomi og forsikring



**Notatnr**  
**Forfattere**

**SAMBA/41/07**  
**Mathilde Wilhelmsen**  
**Xeni K. Dimakos**

**Dato**

**6. november 2007**

## Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Informatikkbygningen ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistikk. Det jobbes med svært mange forskjellige problemstillinger slik som estimering av torskbestand, finansiell risiko, beskrivelse av geologien i petroleumsreservoarer og overvåking av klimaendringer. NR er også ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Problemstillinger kan være å overvåke inntrengning i datasystemer, e-læring i skole og næringsliv, bruk av datateknologi i markedsanalyser samt anvendelser av multimedia på forskjellige plattformer. NRs visjon er forskningsresultater som brukes og synes.

**Tittel** **En studie av sammenhengen mellom makroøkonomi og forsikring**

**Forfattere** **Mathilde Wilhelmsen**  
<Mathilde.Wilhelmsen@nr.no>  
**Xeni K. Dimakos** <Xeni.Dimakos@nr.no>

**Dato** 6. november 2007

**Publikasjonsnummer** SAMBA/41/07

## **Sammendrag**

I dette notatet har vi sett på tidsserier for CR for bolig, landbruk, yrkeskade, uføretet og motorvogn. Vi har forsøkt å se utviklingen i disse i sammenheng med utviklingen i de makroøkonomiske størrelsene KPI, BNP, lønn og arbeidsledighet. Gjensidiges tidsserier er korte og har grov oppløsning, noe som begrenser hvilke analyser som kan gjøres. Med utgangspunkt i dette datagrunnlaget finner vi at det ikke er mulig å trekke noen konklusjoner om sammenhengen mellom utviklingen hos Gjensidige og makroøkonomien.

Emneord	Combined ratio, Korrelasjon, Konjunkturer, Makroøkonomi, Sykler
Målgruppe	Gjensidige Forsikring
Tilgjengelighet	Åpen
Prosjekt	SFI P03 StatInsure
Prosjektnummer	220314
Satsningsområde	SFI
Antall sider	45
© Copyright	Norsk Regnesentral



# Innhold

<b>1</b>	<b>Introduksjon</b>	<b>7</b>
<b>2</b>	<b>Data</b>	<b>7</b>
2.1	Data fra Gjensidige	7
2.2	Makroøkonomiske variable	13
2.3	Glatting og interpolering	18
<b>3</b>	<b>Metode</b>	<b>18</b>
3.1	Autoregressive modeller	18
3.2	Stasjonaritet	19
3.3	Krysskorrelasjon	20
<b>4</b>	<b>Resultater</b>	<b>21</b>
4.1	Gjensidiges tidsrekker	21
4.2	Makroøkonomiske størrelser	22
4.3	Krysskorrelasjoner	22
4.4	Fjerning av trend	22
4.5	Laggede krysskorrelasjoner	33
<b>5</b>	<b>Oppsummering</b>	<b>33</b>
	<b>Appendiks</b>	<b>34</b>
<b>A</b>	<b>Mer metode</b>	<b>34</b>
A.1	Test av stasjonaritet	34
A.2	Fjerning av trend	35

A.3	Fjerning av sesong . . . . .	36
A.4	Kointegrasjon . . . . .	36
	<b>Referanser . . . . .</b>	<b>45</b>

# 1 Introduksjon

I dette notatet ser vi på tidsserier for “combined ratio” (CR, totalkostnadsprosent) for bolig, landbruk, yrkesskade, uførhet og motorvogn. Vi har forsøkt å se utviklingen i disse tidsseriene i sammenheng med utviklingen i de makroøkonomiske størrelsene KPI, BNP, lønn og arbeidsledighet.

Vi ser på hvordan tidsseriene har utviklet seg over tid med hensyn på trend og svingninger (sykler). Videre studerer vi i hvilken grad CR samvarierer med makroøkonomien. Spesielt er vi interessert i om noen av de makroøkonomiske seriene er drivende for utviklingen i Gjensidige. Gjensidiges tidsserier er korte og har grov oppløsning, noe som begrenser hvilke analyser som kan gjøres.

Notatet er organisert som følger. Kapittel 2 presenterer datagrunnlaget. Kapittel 3 beskriver metodene vi benytter, mens kapittel 4 presenterer resultatene. Tilslutt, i kapittel 5, beskriver vi hvilke analyser som kunne vært gjort dersom datagrunnlaget hadde vært større og bedre. Appendikset inneholder metodikk som i såfall er aktuell.

## 2 Data

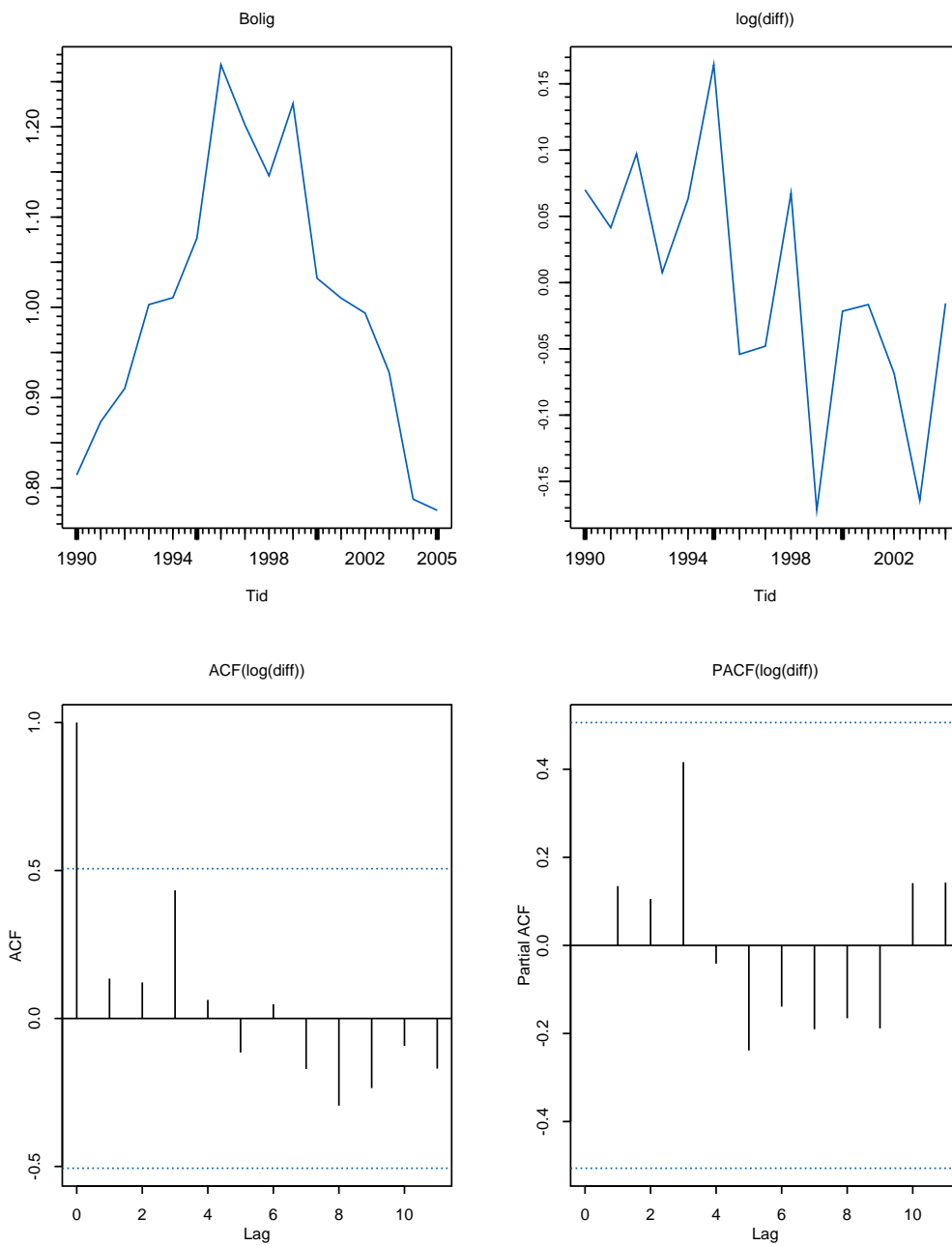
### 2.1 Data fra Gjensidige

Tabell 1 viser en oversikt over skadedata vi har mottatt fra Gjensidige. Alle tidsseriene er på årlig skala.

Skadeprosenten angir hvor stor prosent erstatningskostnaden utgjør av premieinntekten. Kostnadsprosenten angir hvor stor prosent forsikringsrelaterte driftskostnader utgjør av forfalt premie. “Combined ratio” (CR) er summen av skadeprosent og kostnadsprosent, dvs. summen av erstatninger og driftskostnader i forhold til premien. CR blir også kalt totalkostnadsprosent. Dersom CR er større enn 100%, går Gjensidige i underskudd på denne skadetypen.

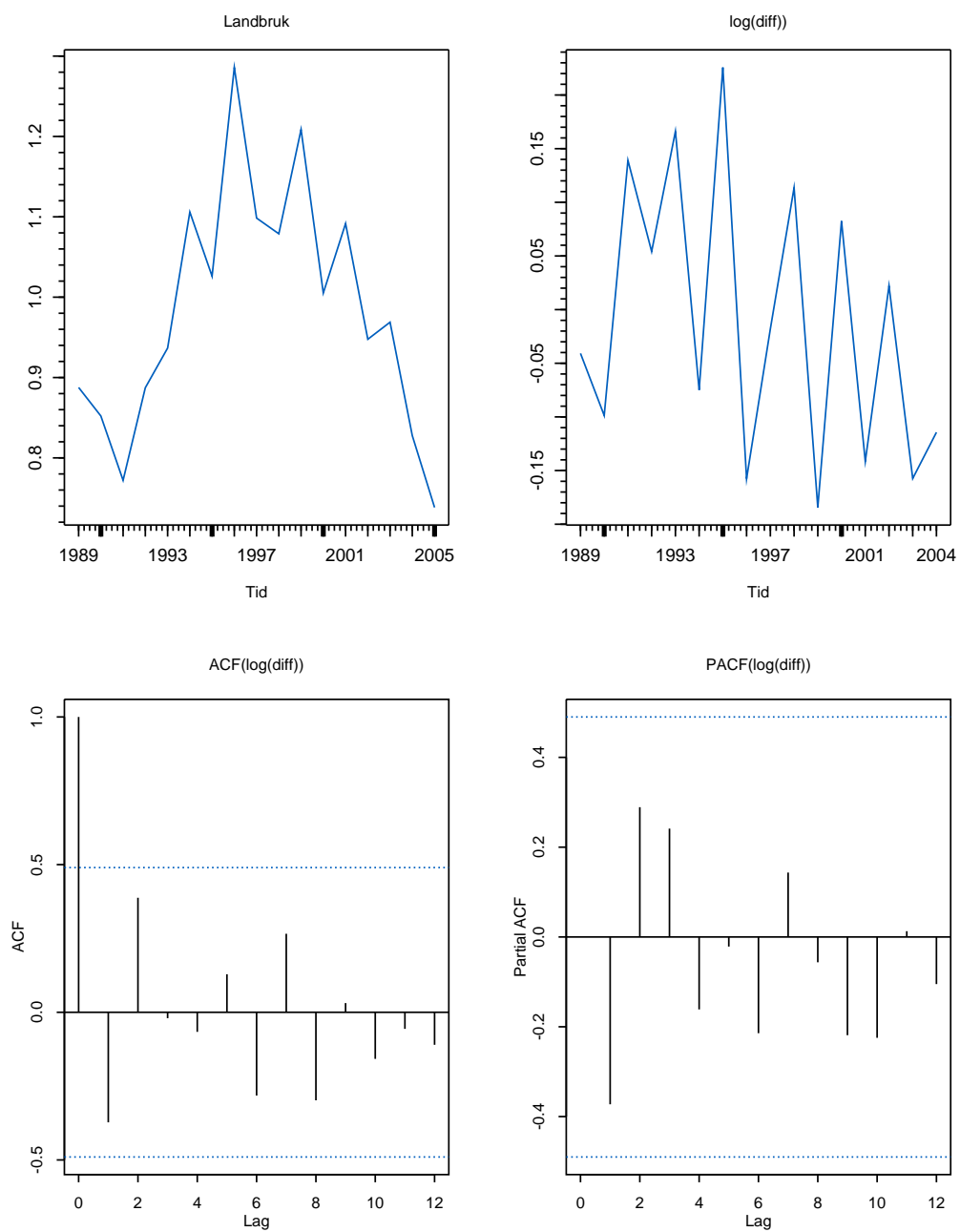
Plottet øverst til venstre i figur 1, 2, 3, 4 og 5 viser CR for de fem skadetyperne fra Gjensidige.

Figurene viser at tidsseriene er for korte til at man kan identifisere flere sykler. For bolig, landbruk, yrkesskade og motorvogn inneholder seriene en periode med oppgang i CR og deretter en periode med nedgang. Mens tidsserien for uføre viser en reduksjon i CR over tid.

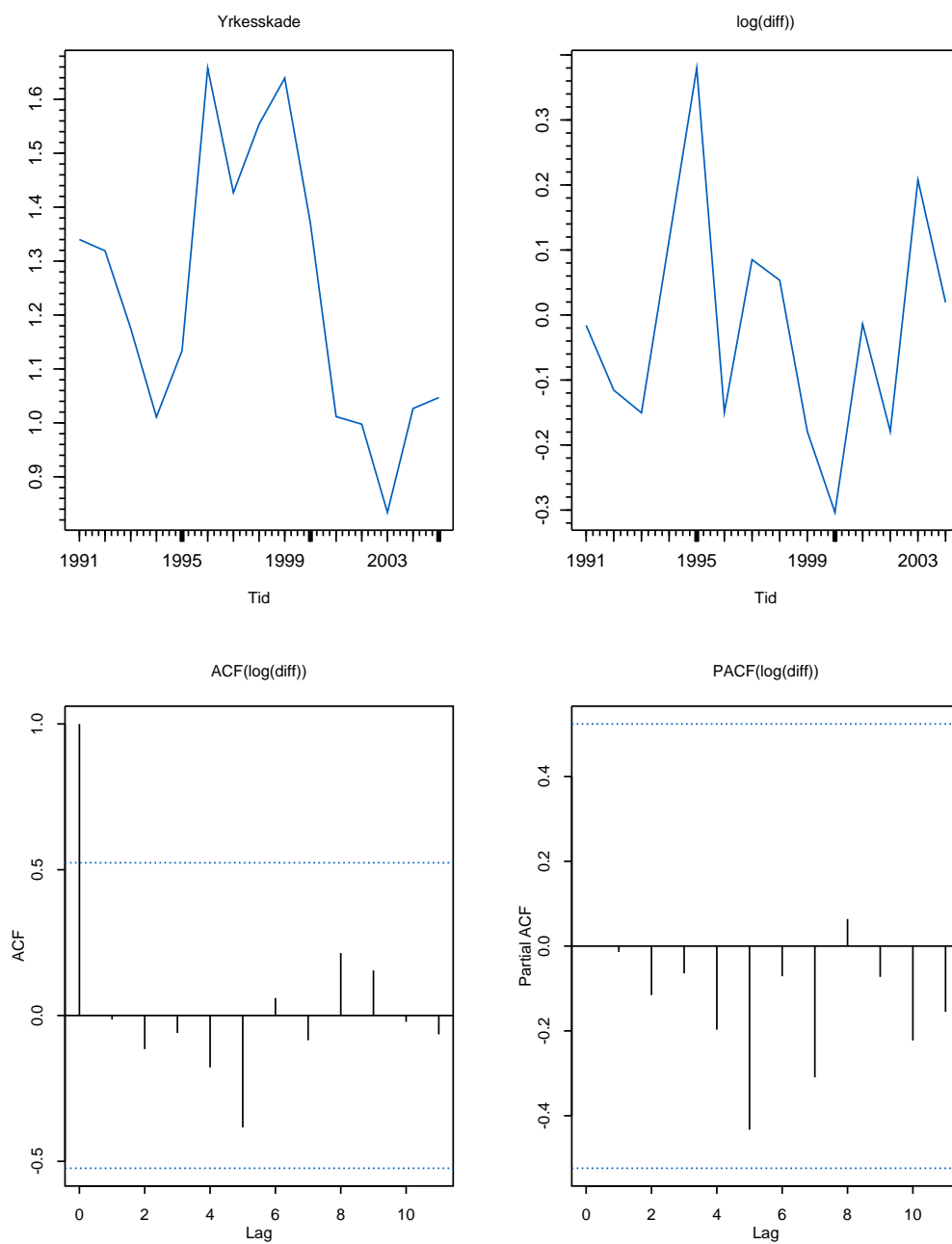


Figur 1. Bolig CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.

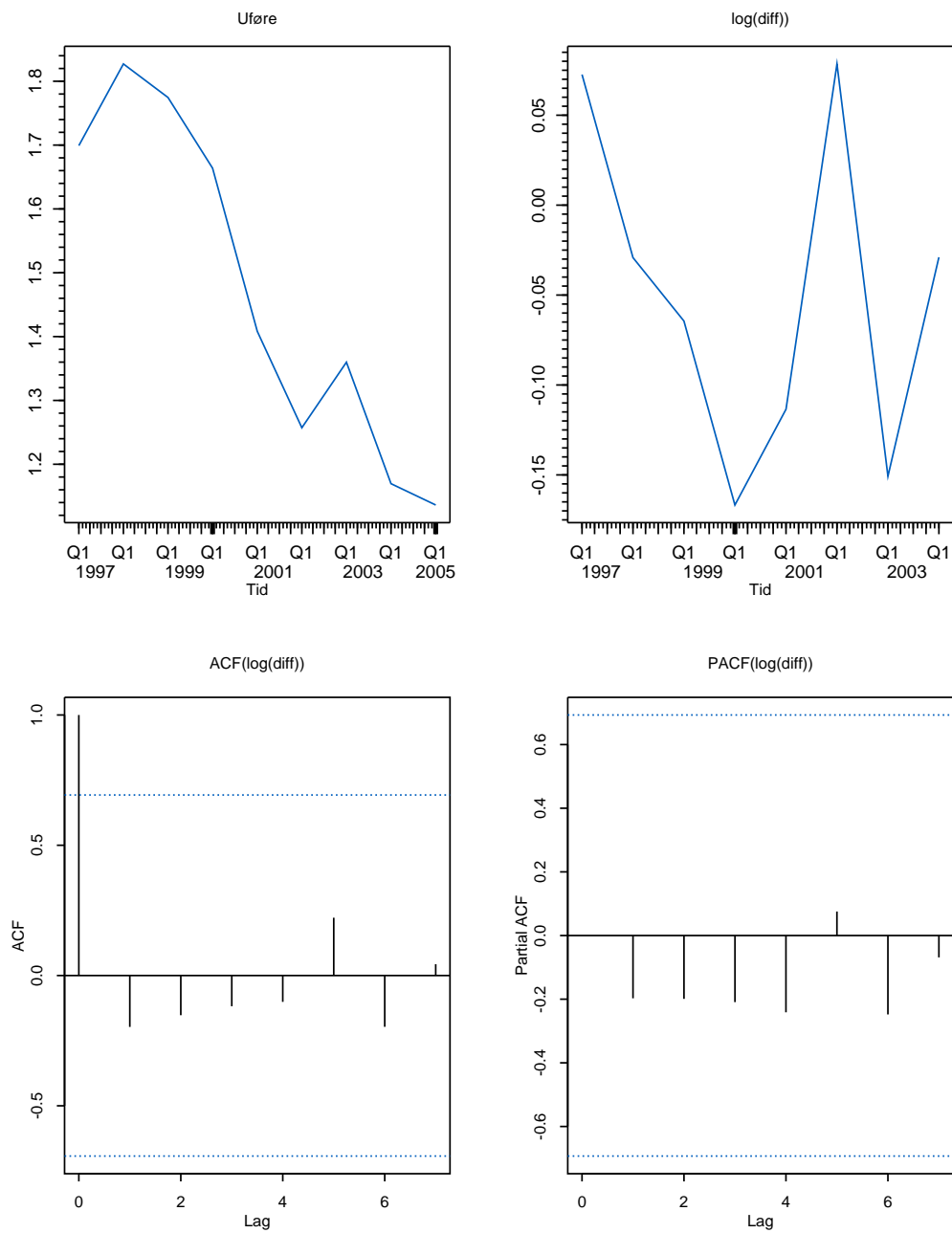




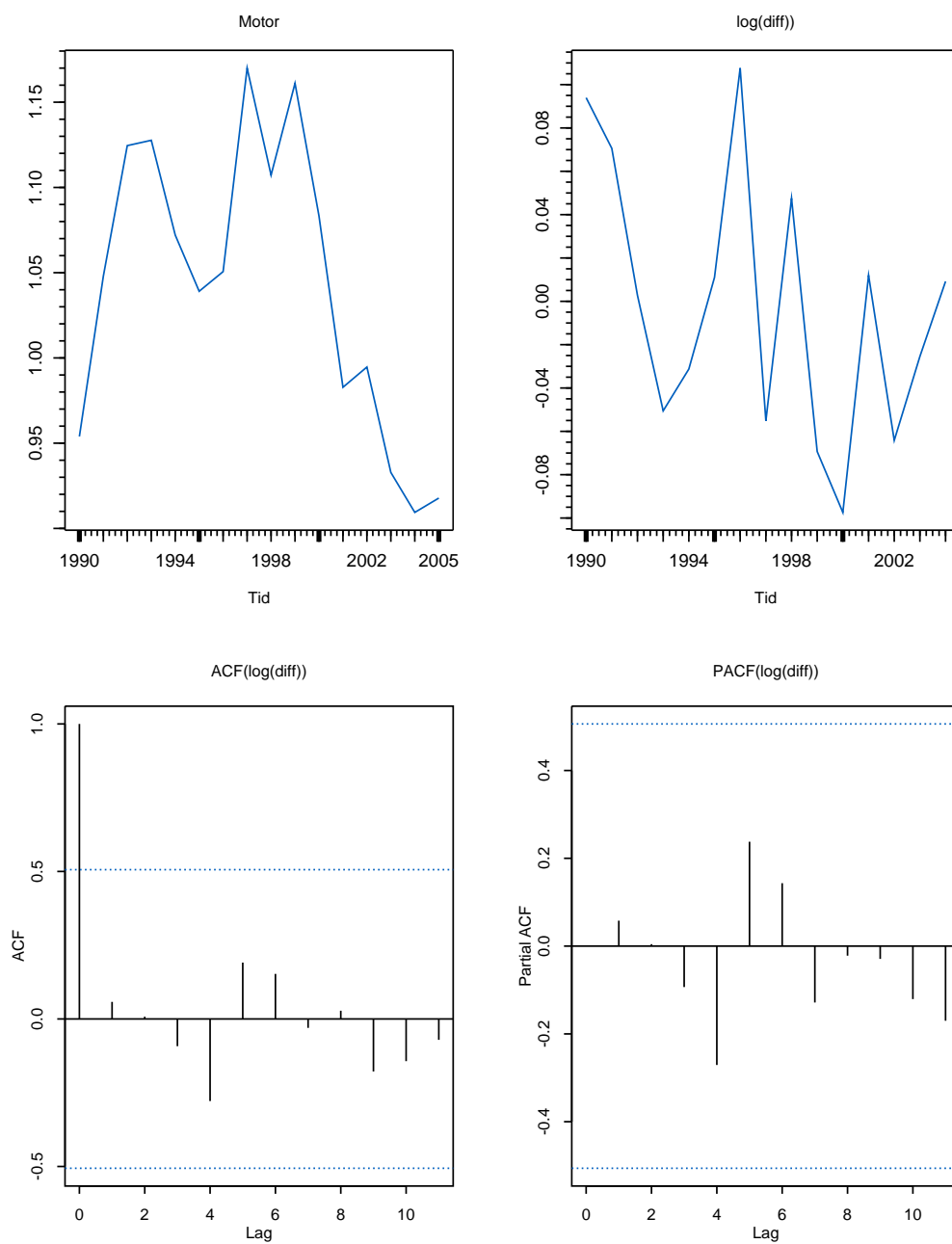
Figur 2. Landbruk CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.



Figur 3. Yrkesskade CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.



Figur 4. Uførhet CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.



Figur 5. Motorvogn CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.

<i>Skadetype</i>	<i>Tilgjengelige størrelser</i>	<i>Oppløsning</i>	<i>Tidsperiode</i>
Bolig	Skadeprocent Kostnadsprosent CR	Årlig	1990-2005
Landbruk	Skadeprocent Kostnadsprosent CR	Årlig	1989-2005
Yrkesskade	Skadeprocent Kostnadsprosent CR	Årlig	1991-2005
Uførhet	Skadeprocent Kostnadsprosent CR Antall skader	Årlig	1997-2005
Motorvogn	Skadeprocent Kostnadsprosent CR	Årlig	1990-2005

Tabell 1. Tabell over skadedata fra Gjensidige

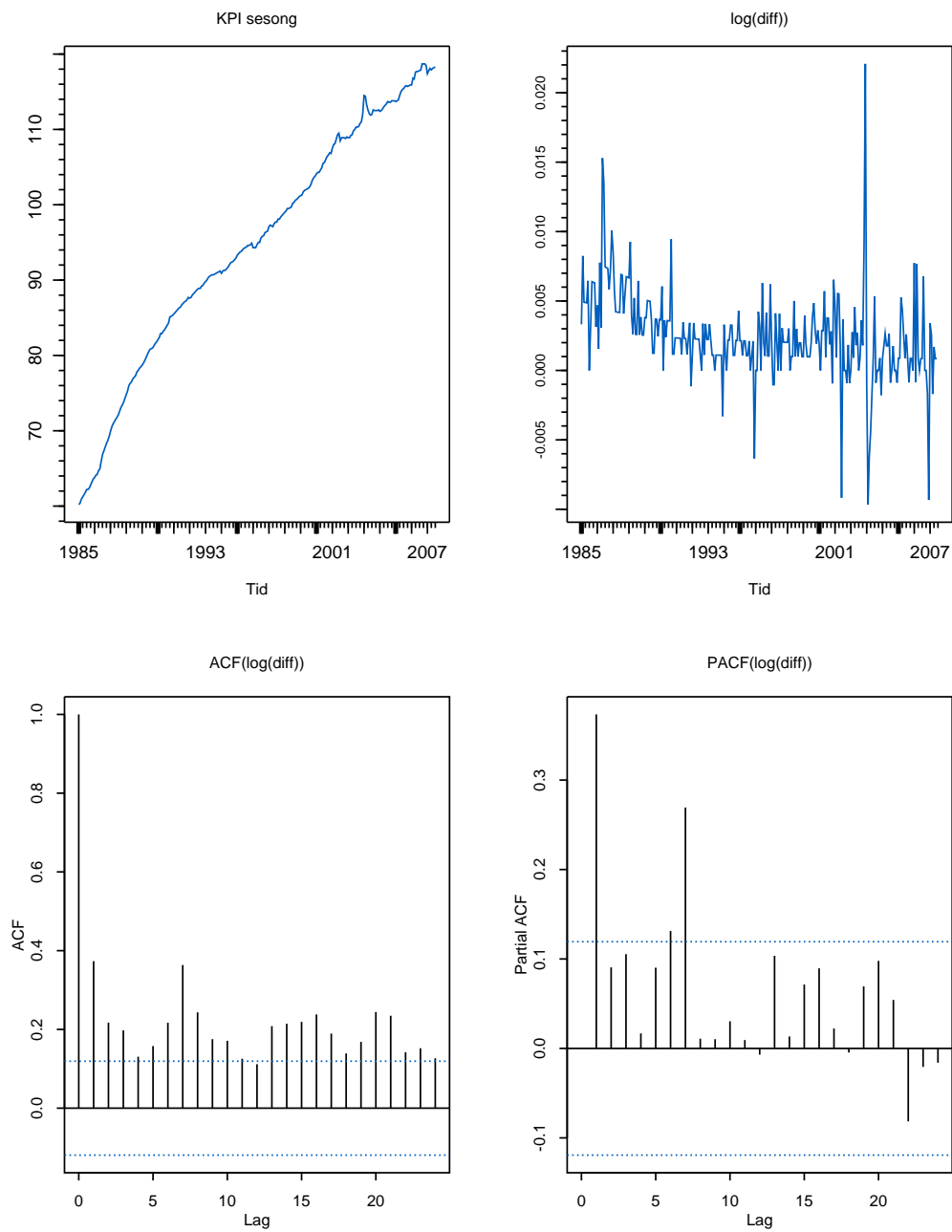
<i>Makroøkonomisk variabel</i>	<i>Oppløsning</i>	<i>Tidsperiode</i>	<i>Kilde</i>
KPI	Månedlig	Januar 1985 - juli 2007	SSB
BNP	Kvartalsvis	1. kvartal 1978 - 2. kvartal 2007	SSB
Lønn	Årlig	1970-2006	SSB
Arbeidsledighetsprosent	Årlig	1954-2005	NAV

Tabell 2. Tabell over makroøkonomiske variable

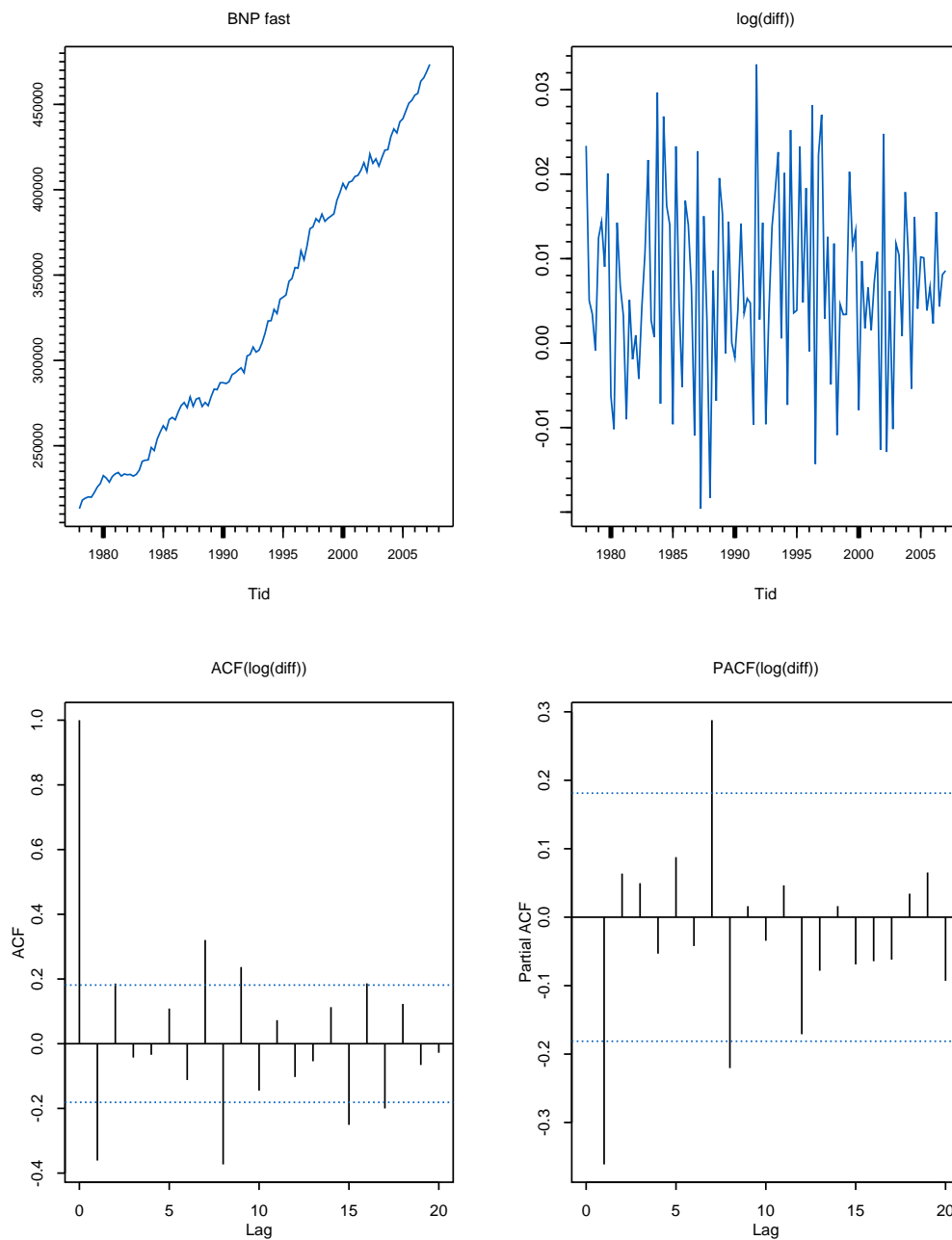
## 2.2 Makroøkonomiske variable

Vi har valgt ut fire makroøkonomiske variable; KPI (Konsumprisindeks), BNP (Brutto nasjonalprodukt), lønn og arbeidsledighetsprosent. Vi bruker en sesongjustert KPI, samt faste verdier for BNP og løpende verdier for lønn. Strengt tatt er det mer korrekt å se på faste verdier, men vi har kun fått tak i løpende verdier for lønn. For våre analyser overskygges dette av at usikkerheten som følger av at Gjensidiges data har så grov tidsoppløsning. Tabell 2 viser en oversikt over variablene.

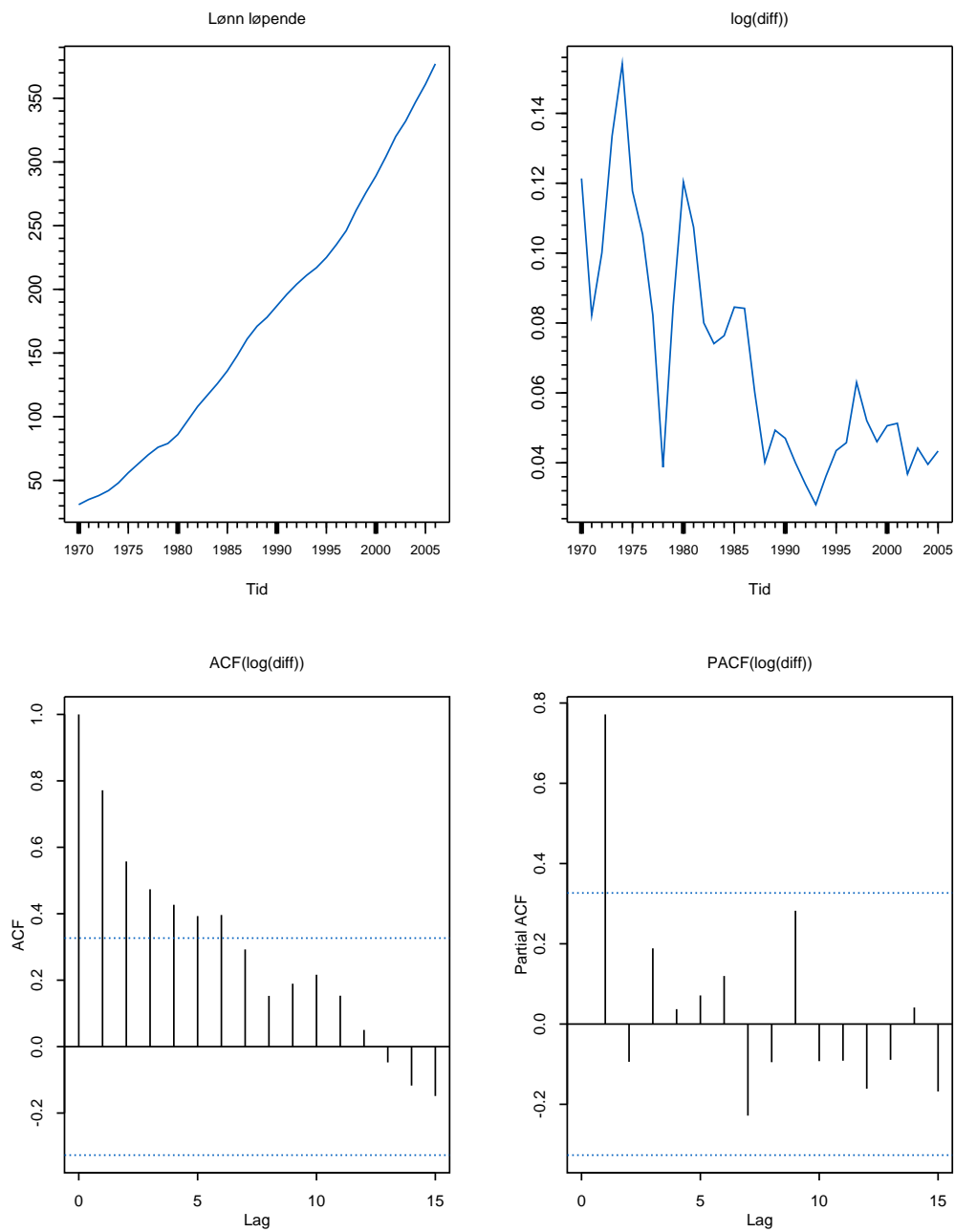
Plottet øverst til venstre i figur 6, 7, 8 og 9 viser tidsseriene for de fire makroøkonomiske variablene. Vi ser at KPI, BNP og lønn ikke overraskende har svært lik utvikling, med en jevn stigning i hele perioden. Arbeidsledigheten er mer varierende, med topper rundt 1985 og 1995, men også perioder med nedgang.



Figur 6. KPI. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.

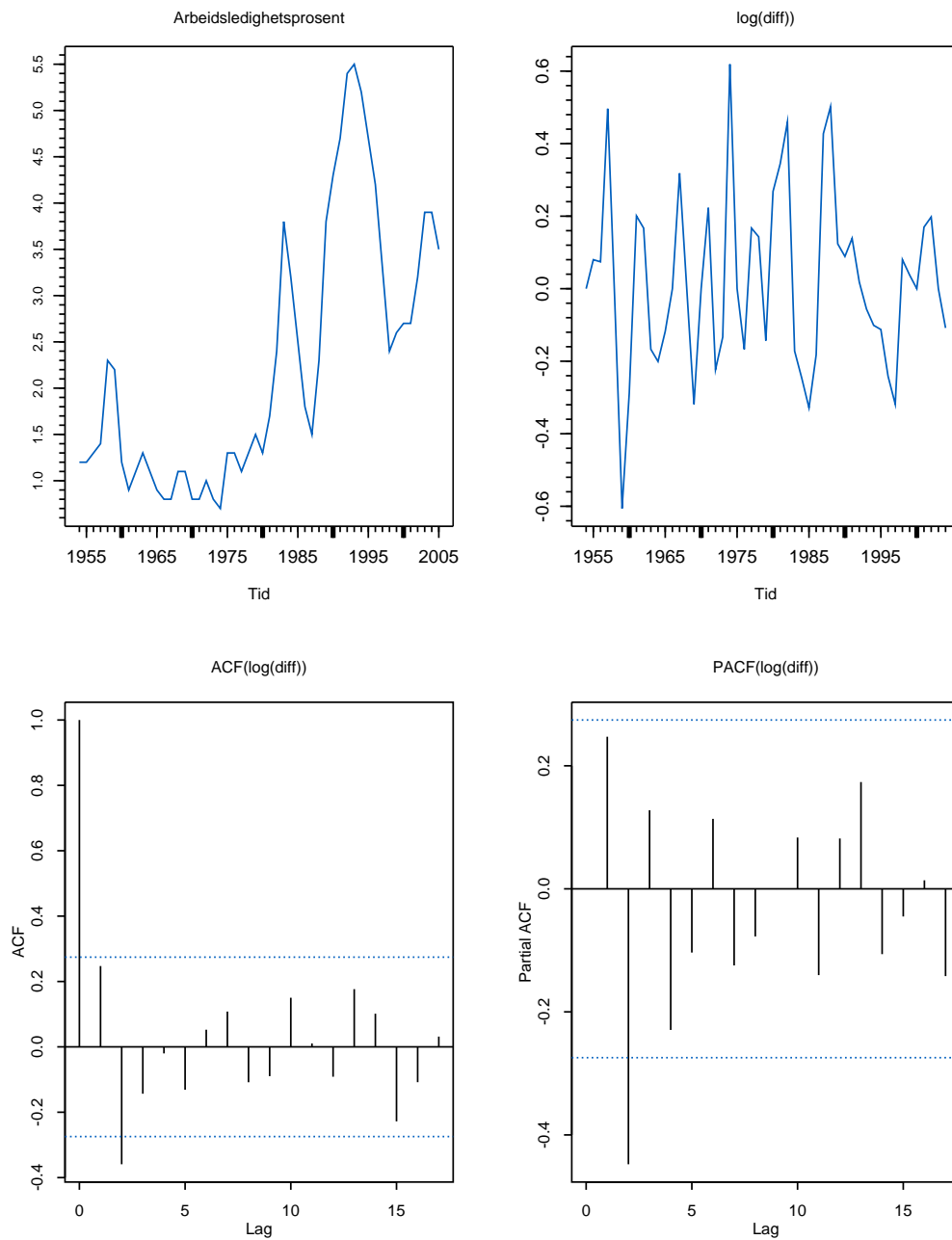


Figur 7. BNP. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.



Figur 8. Lønn. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.





Figur 9. Arbeidsledighetsprosent. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene.

## 2.3 Glatting og interpolering

Når man skal sammenligne to tidsserier, må de ha samme tidsoppløsning. Som vist i de forrige to avsnittene, er ikke dette tilfellet for tidsseriene vi skal studere. Man må da skalere en av tidsseriene slik at den får samme oppløsning som den andre. Vi har valgt å interpolere tidsserien med grovest oppløsning slik at den blir sammenlignbar med den med finest oppløsning. Vi beholder da den fine oppløsningen til den ene tidsserien, og unngår å korte ned lengden på tidsseriene. For å få en best mulig analyse, er det viktig at tidsseriene er lange.

Det er mange måter å interpolere en tidsserie på. Tidsseriene med grovest oppløsning (årlig og kvartalsvis) i de forrige avsnittene består av gjennomsnittsverdier for hver periode. La oss si vi skal sammenligne tidsserie 1 på årlig oppløsning med tidsserie 2 på månedlig oppløsning. Ved å la de årlige gjennomsnittsverdiene være de samme for alle måneder, får vi på en enkel måte gjort om tidsserie 1 til månedlig skala. Denne tidsserien består nå av intervaller med konstante verdier av lengde 12. Det er også mulig å sette de årlige (eller kvartalsvise) gjennomsnittsverdiene til en spesifikk dato (typisk på starten, midten eller slutten av perioden). Splines, samt lineær eller konstant interpolasjon kan så brukes for å "binde" punktene sammen til en tidsserie med høyere oppløsning.

I denne studien har vi gått videre og brukt rutinen `MSmooth` som glatter en slik serie på best mulig måte (Kåresen og Laading, 2000). Denne metoden lager en ny, glatt tidsserie med finere oppløsning enn den originale, ved å bevare gjennomsnittsverdiene i hver periode. Endringene påvirker egenskapene til den originale serien minst mulig.

## 3 Metode

Det kapittelet presenterer metodikken vi benytter i analysene. Før vi forklarer ulike konsepter og metoder, starter vi i avsnitt 3.1 med å beskrive to vanlige måter å representere en tidsserie på.

### 3.1 Autoregressive modeller

En autoregresjon av orden  $p$ , kalt  $AR(p)$ , følger ligningen

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t,$$

der  $\epsilon_t$  er hvit støy, se f.eks. Hamilton (1994).

En ARMA( $p, q$ ) tidsserie er definert som

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q},$$

se igjen Hamilton (1994). Dersom man må differensiere en ARMA( $p, q$ ) tidsserie  $d$  ganger for å få den stasjonær, sier vi at tidsserien er integrert av orden  $d$ , og vi har en ARIMA( $p, d, q$ ) modell. For nærmere beskrivelse av stasjonaritet og integrerte tidsserier, se henholdsvis avsnitt 3.2 og appendiks A.1.

### 3.2 Stasjonaritet

La  $y_t = \{y_t, t = 1, \dots, T\}$  være en tidsserie, dvs. en sekvens av tilfeldige tall med en tidsindeks  $t$ . Tidsserien  $y_t$  er *kovariansstasjonær* hvis

$$E[y_t] = \mu \quad \text{for alle } t$$

og

$$\text{Cov}(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \quad \text{for alle } t \text{ og } j,$$

der  $E$  og  $\text{Cov}$  står for henholdsvis forventning og kovarians. Vi vil for enkelthets skyld kalle en kovariansstasjonær tidsserie for en stasjonær tidsserie. Parametere  $\gamma_j$  er *autokovariansen* til  $y_t$  med orden (lag)  $j$ . Et plott av  $\gamma_j$  for alle  $j$  blir kalt autokovariansfunksjonen. Autokorrelasjonene til  $y_t$  er definert som

$$\rho_j = \frac{\text{Cov}(y_t, y_{t-j})}{\sqrt{\text{Var}(y_t)\text{Var}(y_{t-j})}} = \frac{\gamma_j}{\gamma_0},$$

der variansen  $\text{Var}(y_t) = \text{Cov}(y_t, y_t) = \gamma_0$  er den samme for alle  $t$  som følge av stasjonariteten. Et plott av  $\rho_j$  for alle  $j$  blir kalt autokorrelasjonsfunksjonen (ACF). Et naturlig estimat for autokorrelasjonen  $\rho_j$  er gitt ved  $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$ , der

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y}) \quad \text{for } j = 0, 1, \dots, T-1,$$

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t.$$

En annen nyttig størrelse er den *partielle autokorrelasjonsfunksjonen* (PACF)

$$a_{k,k} = \frac{\hat{\gamma}_k - \sum_{i=1}^{k-1} a_{i,k-1} \hat{\gamma}_{k-i}}{\sigma_{k-1}^2},$$

der

$$a_{i,k} = a_{i,k-1} - a_{k,k} a_{k-i,k-1} \quad \text{for } 1 \leq i \leq k-1$$

er et estimat av den  $i$ te autoregresjonskoeffisienten  $\phi_i$  i en  $AR(k)$  modell. Videre er

$$\sigma_k^2 = \sigma_{k-1}^2(1 - a_{k,k}^2).$$

Siden det kan vises at  $AR(p)$  koeffisientene  $\phi_1, \phi_2, \dots, \phi_p$  tilfredsstillter *Yule-Walker* ligningene

$$\gamma_m = \sum_{r=1}^p \phi_r \gamma_{m-r} \quad \text{for } m = 1, 2, \dots, p,$$

kan kvadratet av  $a_{k,k}$  tolkes som et mål på nytten av å øke ordenen på  $AR$  prosessen fra  $k - 1$  til  $k$ . Sekvensen  $a_{k,k}$ ,  $k = 1, \dots$ , kan brukes for å sjekke om en serie er en  $AR$  prosess. Dersom serien er en  $AR(p)$  prosess, bør alle  $a_{k,k}$  være nær null for  $k > p$ . Se for eksempel Hamilton (1994) og MathSoft (2000) for mer detaljer.

### 3.3 Krysskorrelasjon

Korrelasjon er det enkleste målet på samvariasjon mellom to variable. Dersom man ser på tidsrekker, kalles ofte korrelasjonene mellom to tidsrekker for krysskorrelasjon. Dette er for å skille begrepet fra autokorrelasjon, som angir korrelasjonen mellom ulike tidspunkt i samme serie.

For to tidsrekker  $\mathbf{x}_t = \{x_t, t = 1, \dots, T\}$  og  $\mathbf{y}_t = \{y_t, t = 1, \dots, T\}$  definerer vi (kryss)korrelasjonen som

$$\text{Corr}(x_t, y_t) = \text{Cov}(x_t, y_t) / \sqrt{\text{Var}(x_t)\text{Var}(y_t)},$$

som estimeres ved den empiriske korrelasjonskoeffisienten

$$\widehat{\text{Corr}}(x_t, y_t) = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{s_x s_y},$$

der  $s_x$  og  $s_y$  angir empirisk standardavvik for tidsrekkene  $\mathbf{x}_t$  og  $\mathbf{y}_t$ .

Dersom krysskorrelasjonen mellom to tidsrekker er positiv, kaller vi de to tidsrekkene *pro-sykliske*. Dersom den er negativ, sier vi at seriene er *mot-sykliske*, mens dersom korrelasjonen er nær 0, sier vi at de er *asykliske*. Denne terminologien er hentet fra Bjørnland (2000) som har gjort slike studier på restleddene etter at trend og sesong er fjernet.

Korrelasjonene kan også studeres på ulike tidspunkt i de to tidsrekkene, dvs. når tidsrekkene er forskjøvet i forhold til hverandre. Hvis den høyeste korrelasjonen mellom en tidsrekke fra Gjensidige og en makroøkonomisk størrelse opptrer dersom serien forskyves forover i forhold til den makroøkonomiske størrelsen, sier

vi at den makroøkonomiske variabelen er *ledende* for Gjensidiges variabel. Dersom vi lar  $m_t$  angi den makroøkonomiske størrelsen og  $g_t$  Gjensidiges størrelse tilsvarer dette at  $\text{Corr}(m_t, g_{t+k})$  er større for en positiv  $k$ , enn for  $k = 0$ . Negative verdier for  $k$  tilsvarer at Gjensidiges variabel er ledende for den makroøkonomiske variabelen, hvilket er svært lite sannsynlig.

Studier av *rullerende* korrelasjoner viser hvordan korrelasjonen varierer over tid ved at man beregner korrelasjoner for et tidsvindu av en viss lengde og flytter dette vinduet stegvis fremover.

Vi vil studere i hvilken grad tidsseriene fra Gjensidige samvarierer med de makroøkonomiske størrelsene. I slike sammenligninger bør man benytte sesongjusterte data dersom det er mulig. Når det gjelder tidsseriene fra Gjensidige, så har disse årlig oppløsning og sesongjusteringer er derfor ikke nødvendig. For SSB-dataene har vi benyttet sesongjusterte verdier der de har vært tilgjengelige. Dette samsvarer med hvilke variable SSB mener det er fornuftig å sesongjustere. Vi vil både se på samvariasjon av de originale tidsseriene og etter at trenden er fjernet fra seriene. Ved å fjerne trenden ser vi på hvordan utviklingen i tidsseriene utover trenden samvarierer.

## 4 Resultater

### 4.1 Gjensidiges tidsrekker

Figur 1, 2, 3, 4 og 5 viser CR for de fem skadetyperne fra Gjensidige. Plottene i øverste venstre hjørne viser selve tidsrekken. De fleste skadetyperne har en periode med økning i CR og en periode med nedgang. Tidsrekkene inneholder altså maksimalt én sykel, noe som medfører at det ikke er mulig å etablere statistiske modeller for syklene.

Plottene i øverste høyre hjørne viser differansen mellom logaritmen av tidsrekken på et tidspunkt og tidspunktet før, dvs.  $\log y_t - \log y_{t-1}$ , og viser at det er relativt store variasjoner fra et tidspunkt til et annet på log-skala.

De nederste plottene viser autokorrelasjonsfunksjonen og den partielle autokorrelasjonsfunksjonen for logdifferansene. Autokorrelasjonsfunksjonen viser at endringene i CR påfølgende år er svært lite korrelerte, noe som ikke er overraskende med en så grov oppløsning på dataene. De partielle autokorrelasjonene viser ikke noe entydig mønster og indikerer at usikkerheten er for stor til at man kan slutte noe om hvorvidt tidsrekkene kan beskrives som autoregressive prosesser.

## 4.2 Makroøkonomiske størrelser

De makroøkonomiske størrelsene KPI, BNP, lønn og arbeidsledighet er plottet i hhv. figur 6, 7, 8 og 9. Figurene viser at KPI, BNP og lønn har hatt en jevn stigning, mens arbeidsledigheten har hatt både kraftige økninger og nedganger i perioden. Figurene viser at log-differansen for KPI, BNP og arbeidsledighet synes å være stasjonære prosesser, mens stasjonariteten for lønn er noe mer usikker. Autokorrelasjonsfunksjonene for KPI og lønn viser en positiv autokorrelasjon mellom verdier nær i tid, mens BNP har en negativ autokorrelasjon mellom påfølgende tidspunkter. Også for de makroøkonomiske variablene er de partielle autokorrelasjonene beheftet med en så stor usikkerhet at det ikke er mulig å trekke noen konklusjoner.

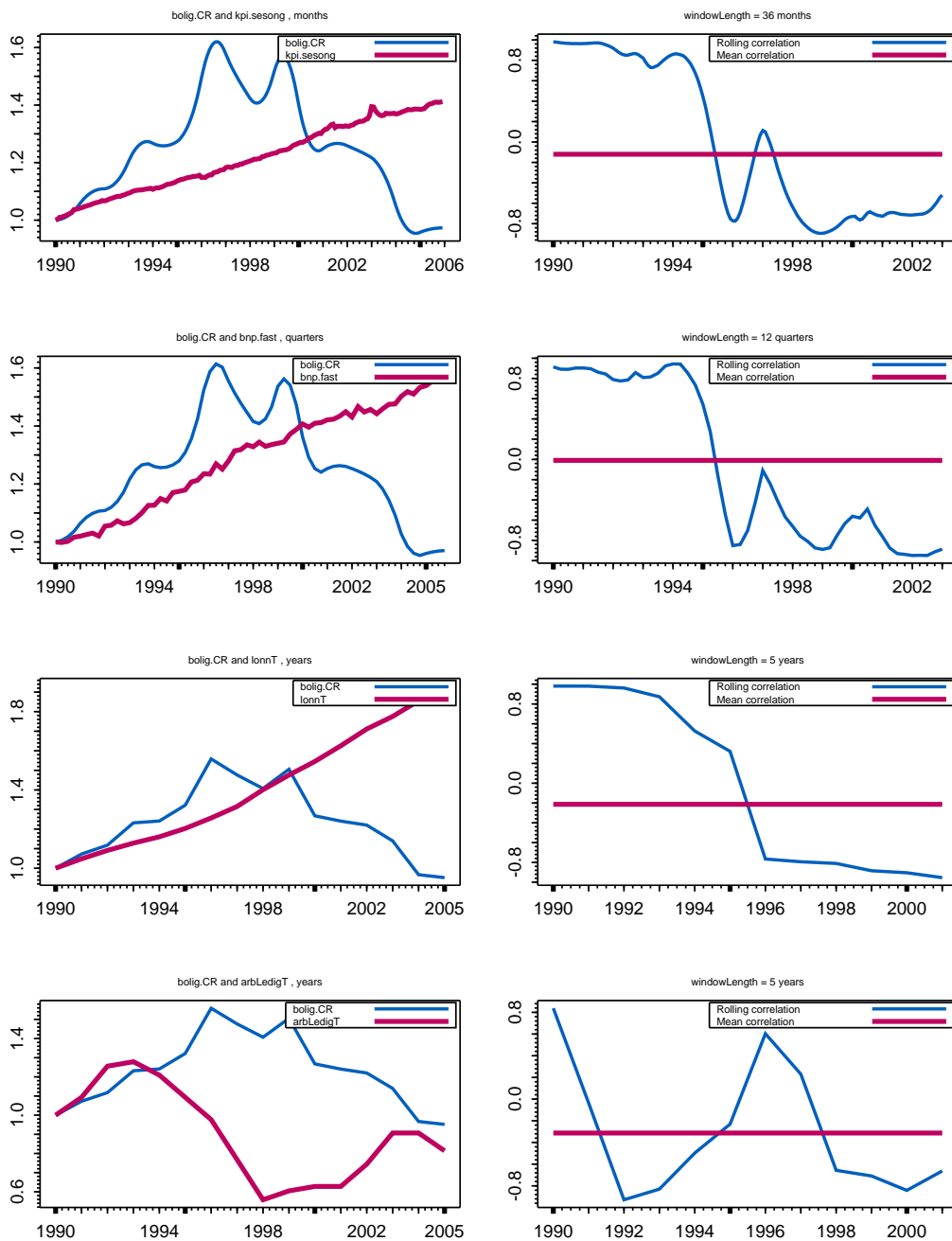
## 4.3 Krysskorrelasjoner

Figur 10, 11, 12, 13 og 14 viser hver av tidsseriene til Gjensidige plottet mot hver av de makroøkonomiske variablene, samt tilhørende rullerende krysskorrelasjoner, uten at seriene er forskjøvet i forhold til hverandre. Tidsvinduene som er brukt ved beregning av rullerende krysskorrelasjoner varierer fra serie til serie, men det er typisk brukt et vindu på 3 – 5 år. Som det framgår av figurene, er det vanskelig å se et klart mønster i de rullerende krysskorrelasjonene. For de av Gjensidiges serier som består av en oppgang og en nedgang, blir korrelasjonen med makrovariablene nær null. Den rullerende korrelasjonen viser dette ved at den først er positiv (i oppgangsperioden) og deretter negativ (i nedgangsperioden). For uførhet, som har en nedadgående trend i hele perioden, blir korrelasjonen negativ.

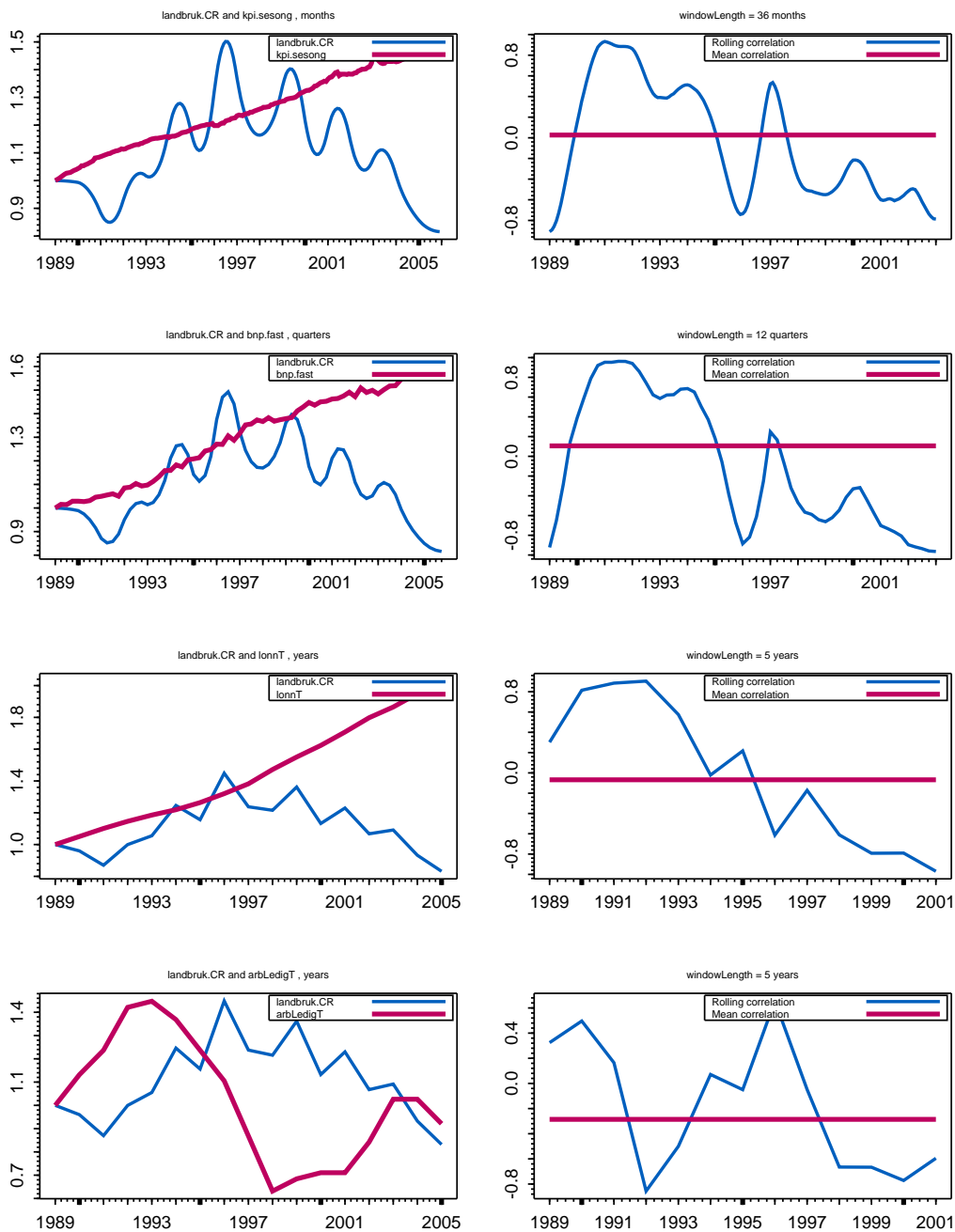
Endringene i Gjensidige CR data ble i tillegg sammenlignet med endringene i de makroøkonomiske dataene, men heller ikke disse viste et tydelig mønster i de rullerende krysskorrelasjonene.

## 4.4 Fjerning av trend

Vi har også forsøkt å fjerne eventuelle trender i tidsseriene, og deretter sammenlignet de gjestående residualene med hverandre. Den innebygde Splus funksjonen `loess.smooth` gjør en lokal polynomregresjon av dataene, som vi lar være trenden. Figur 15, 16, 17, 18 og 19 viser hver av CR tidsseriene (uten trend) plottet mot hver av de makroøkonomiske variablene (uten trend), samt de tilhørende krysskorrelasjonene, også her uten lag. Heller ikke nå er det lett å se et klart mønster i samvariasjonen mellom variablene.

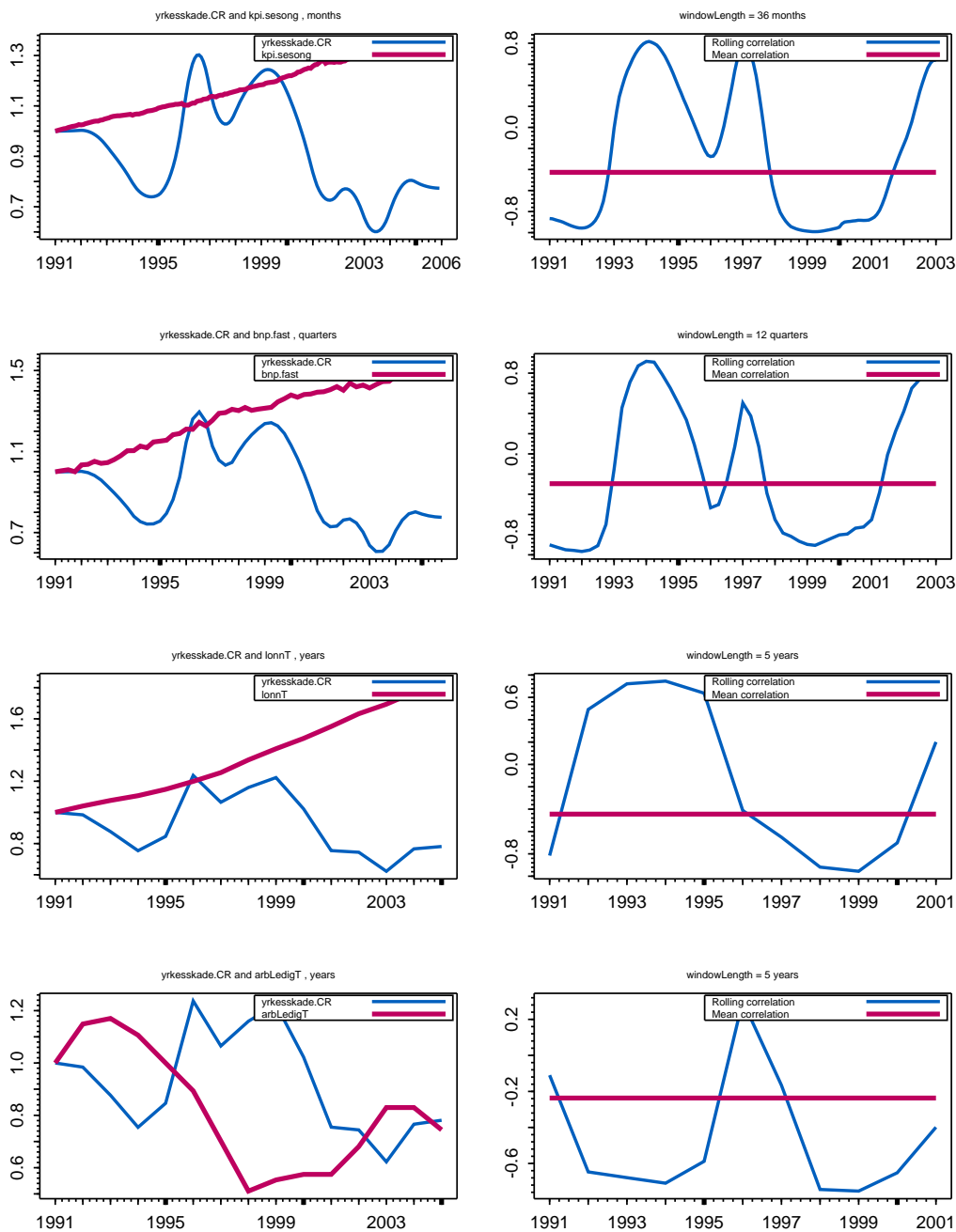


Figur 10. Bolig CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Bolig CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.

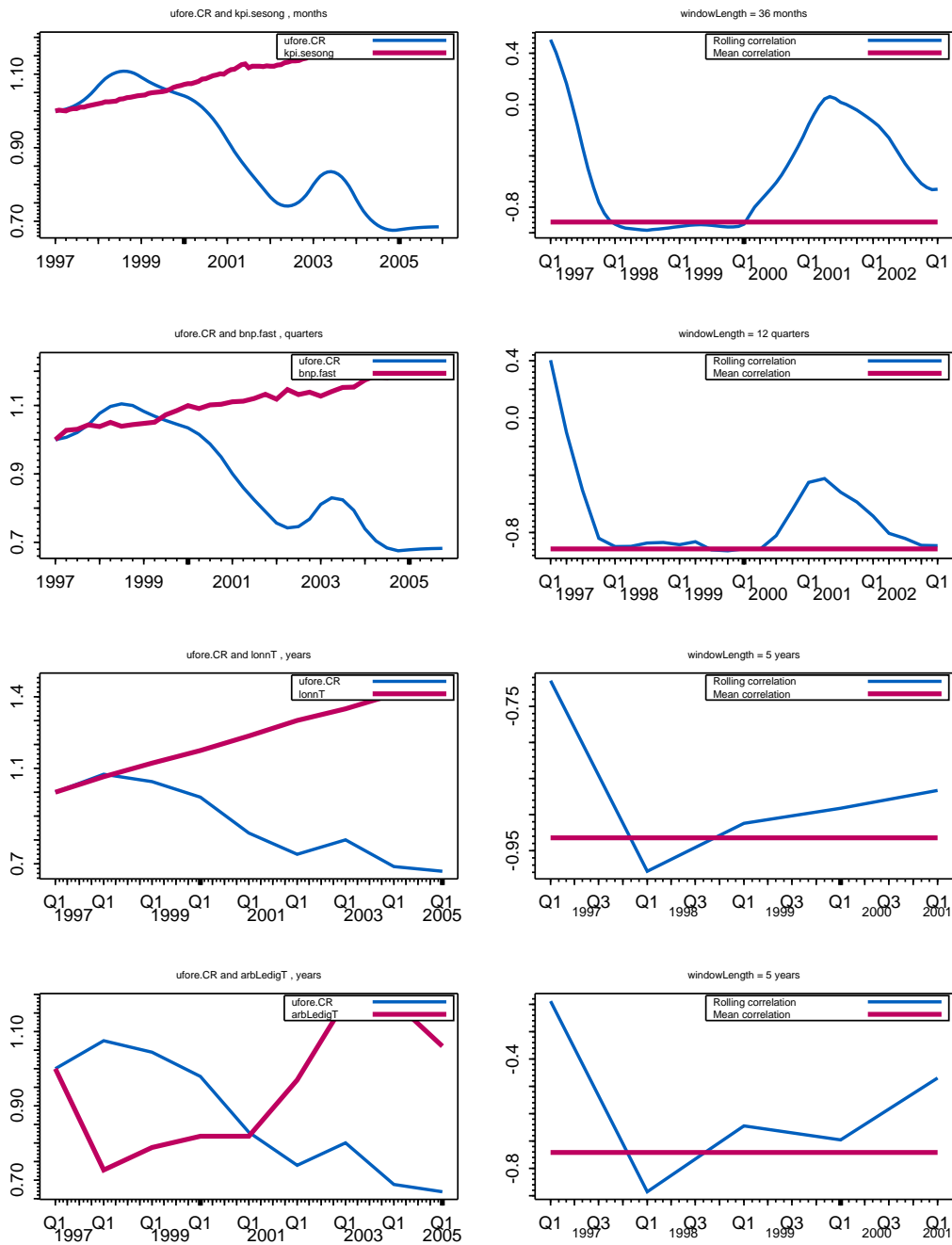


Figur 11. Landbruk CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Landbruk CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.

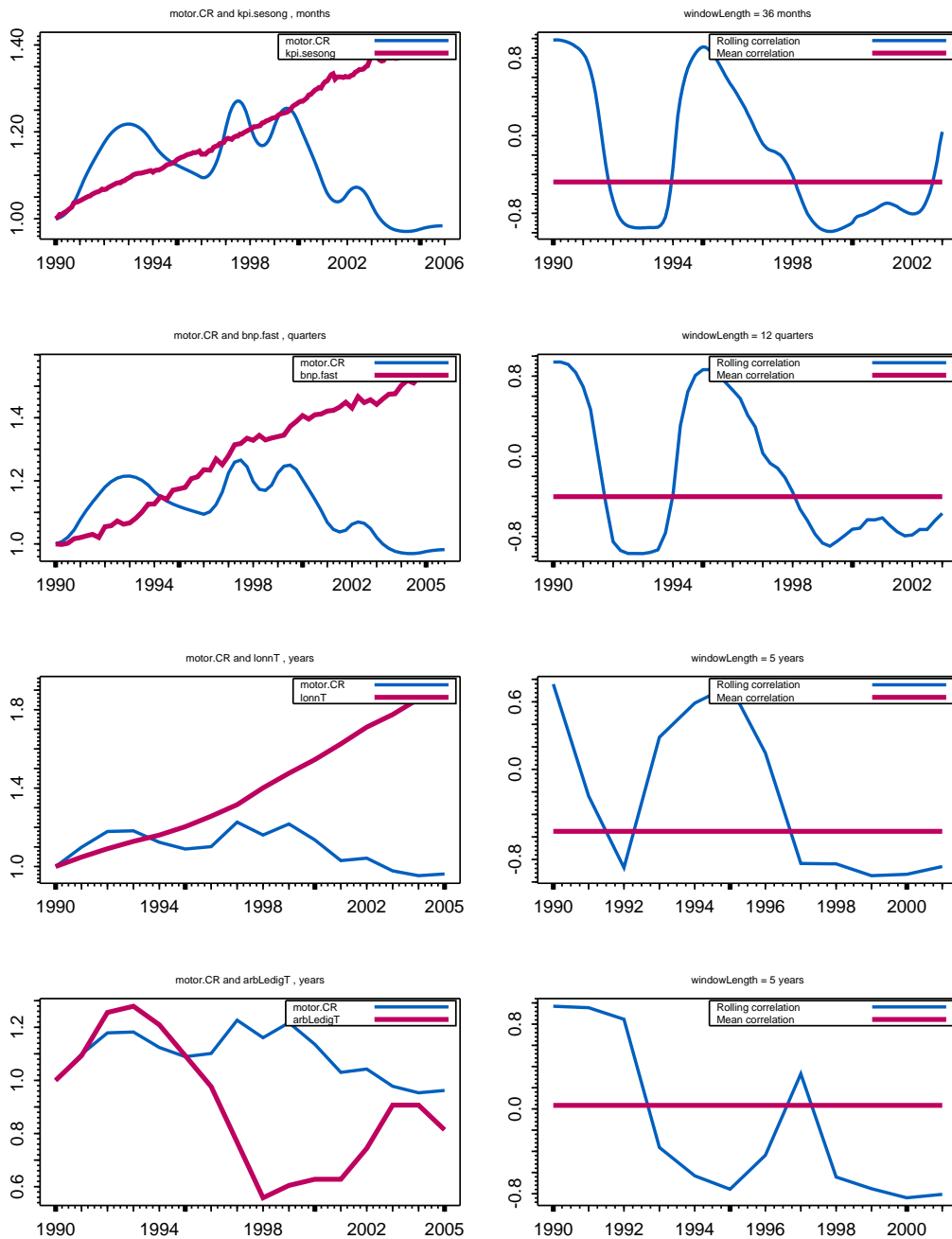




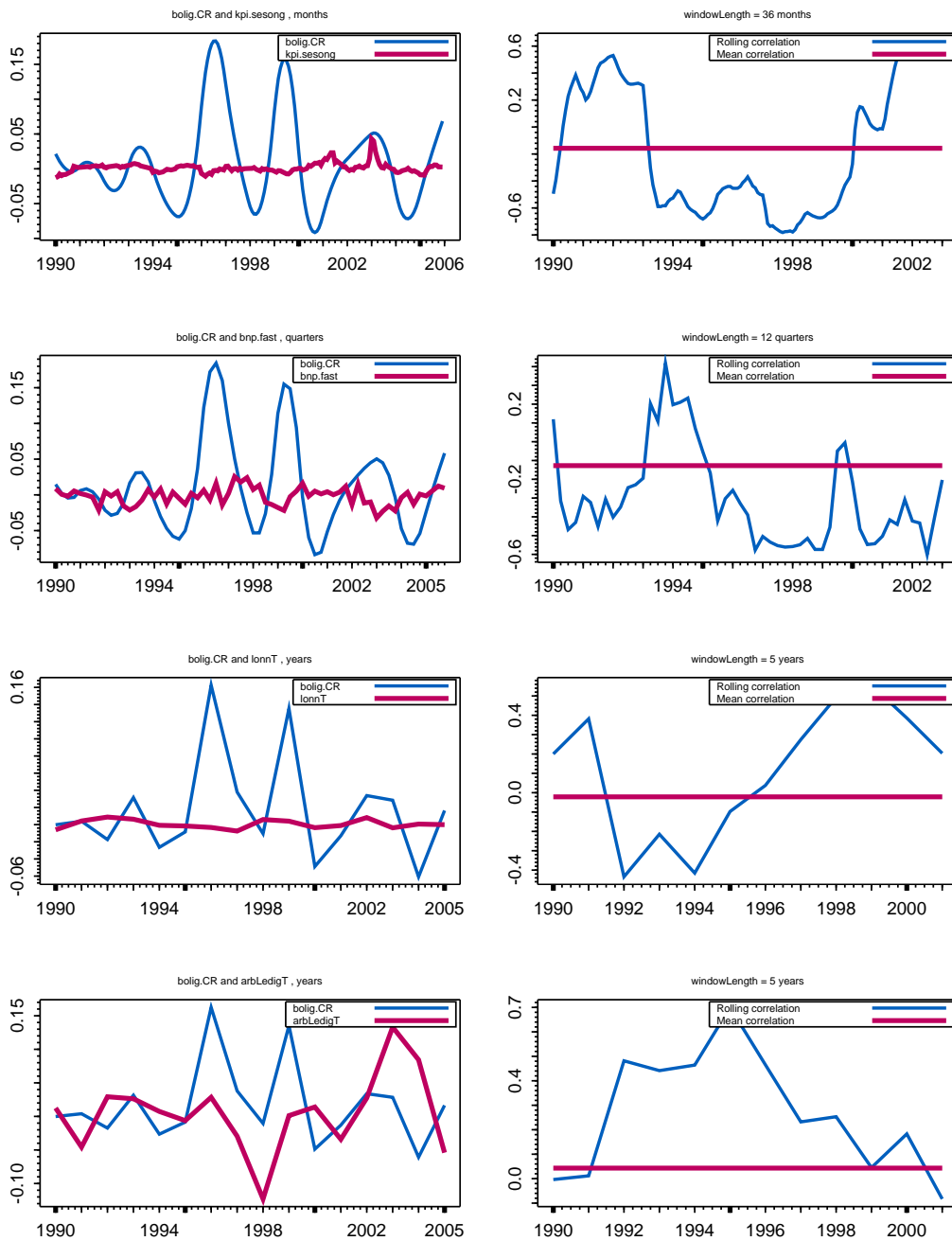
Figur 12. Yrkesskade CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Yrkesskade CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



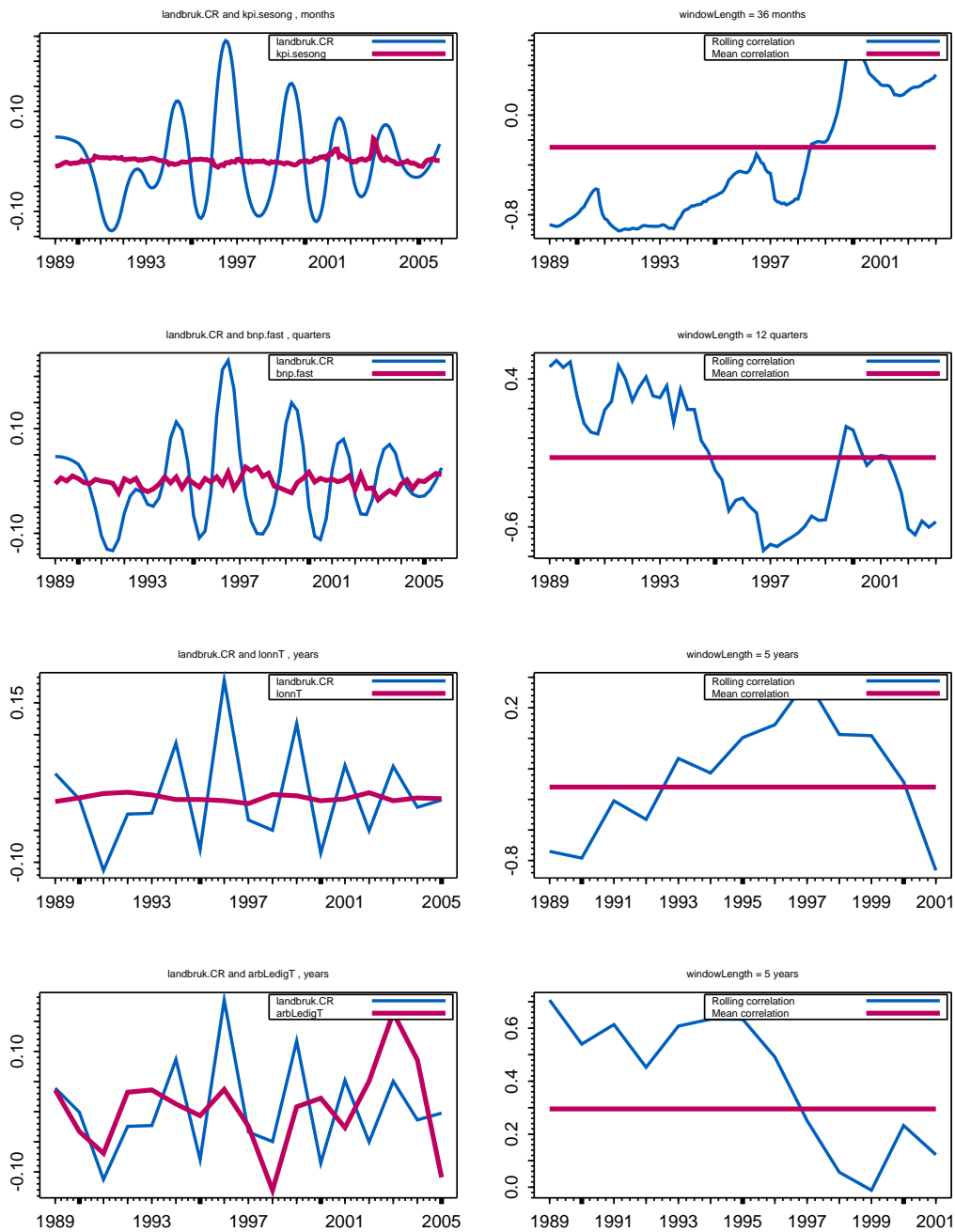
Figur 13. Uføre CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Uføre CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



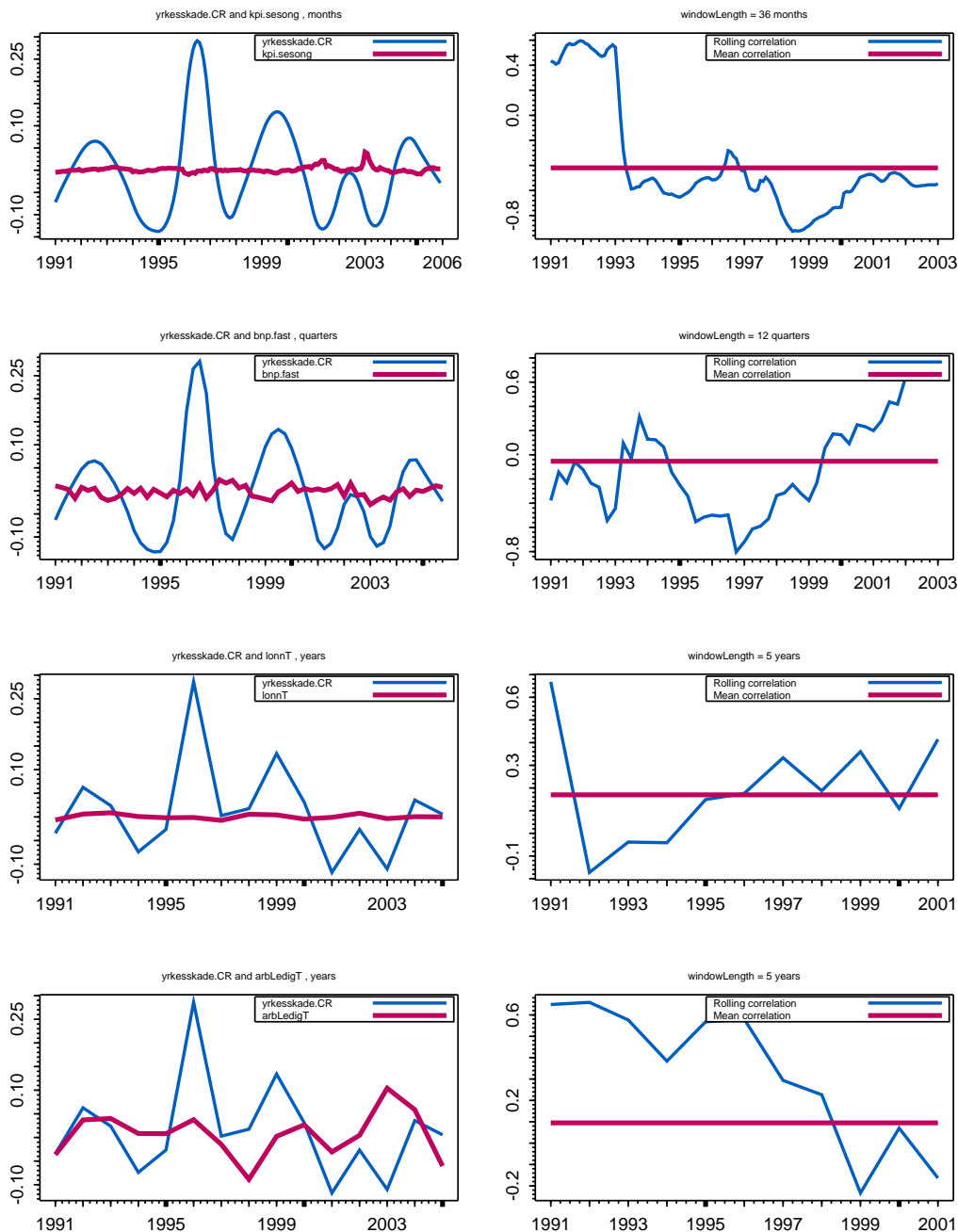
Figur 14. Motor CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Motor CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



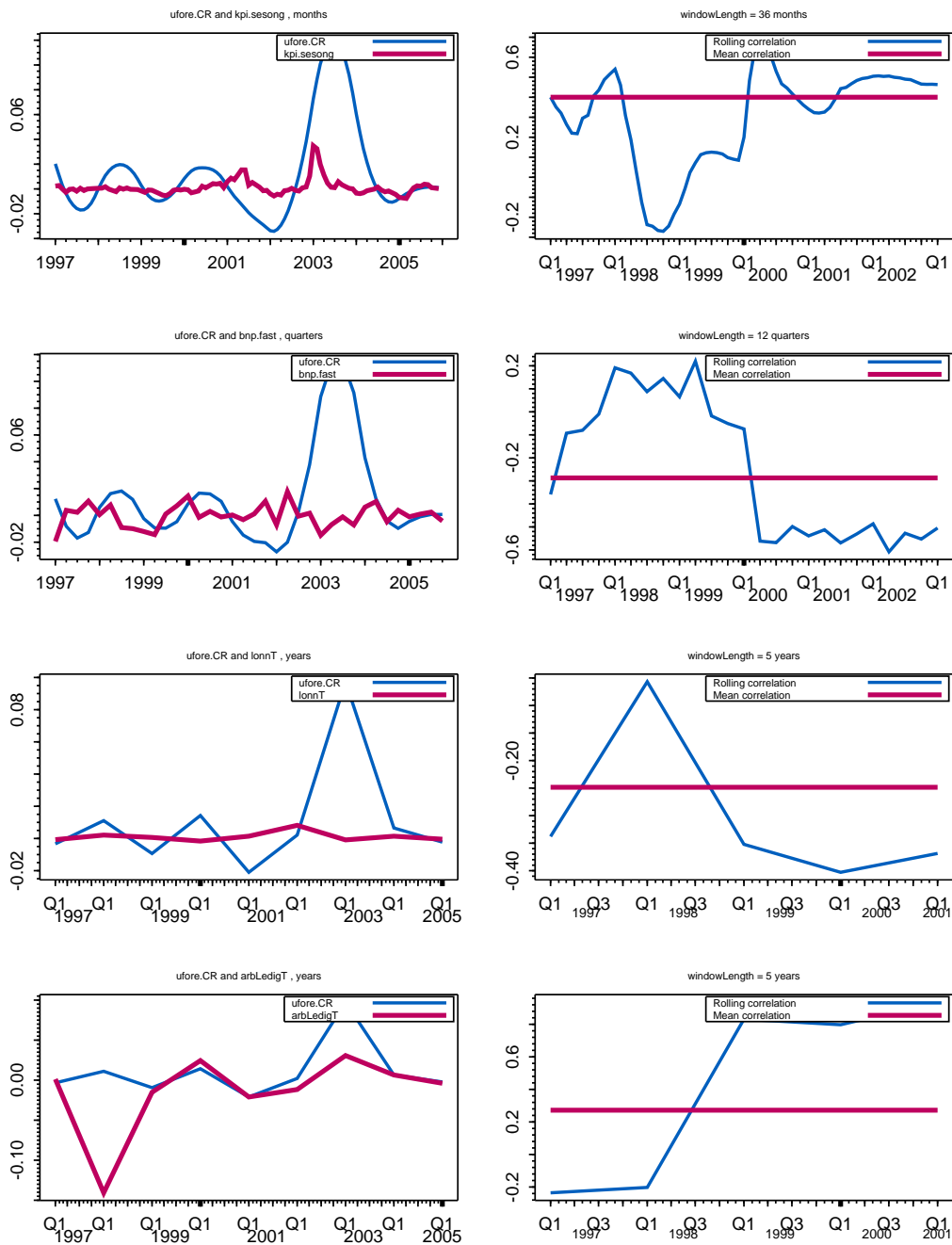
Figur 15. Bolig CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Bolig CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



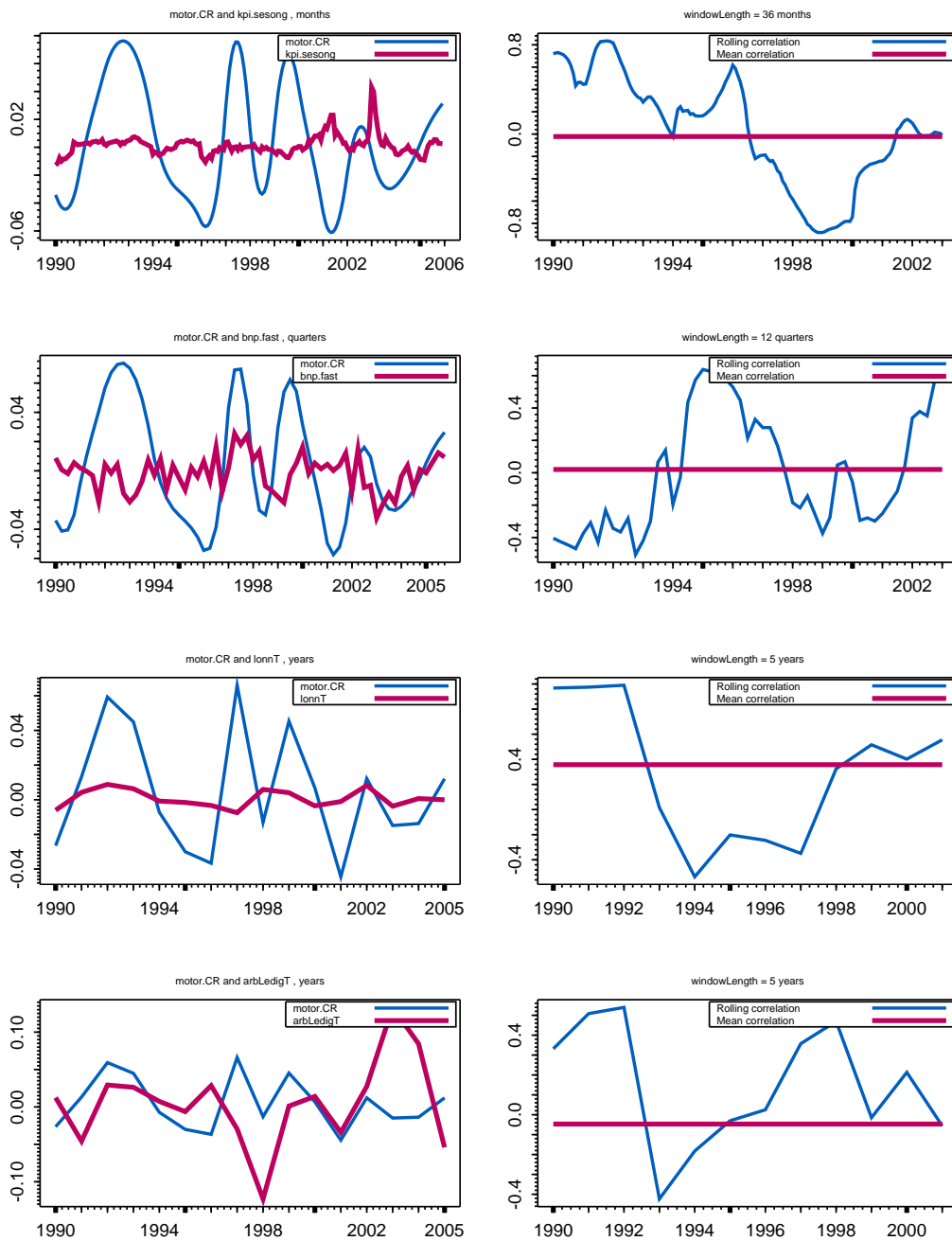
Figur 16. Landbruk CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Landbruk CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



Figur 17. Yrkesskade CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Yrkesskade CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



Figur 18. Uføre CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Uføre CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



Figur 19. Motor CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Motor CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre.



## 4.5 Laggede krysskorrelasjoner

Vi forsøkte å se om det skjedde systematiske endringer i korrelasjonene dersom vi forskjøv tidsseriene i forhold til hverandre. Utgangspunktet var altså at vi ønsket å se om de makroøkonomiske variablene var ledende for Gjensidiges størrelser. Vi undersøkte en forskyvning på et og to år. For bolig, landbruk, yrkesskade og motorvogn, som alle har tidsserier som består av en oppgangsperiode og en nedgangsperiode, øker korrelasjonen mot KPI, BNP og lønn. Alle disse makrovariablene stiger jevnt i perioden vi betrakter. Denne økningen er et resultat av at forskyvningen gjør at oppgangsperioden ligger nærmere utviklingen i makrovariablene. For arbeidsledigheten synker korrelasjonen når forskyvningen gjøres. For uførhet har forskyvningen ingen betydning. De rullerende korrelasjonene viser seg fortsatt å variere sterkt i perioden. Dette betyr at vi ikke kan konkludere med at makrovariablene er ledende, selv om noen av de laggede gjennomsnittskorrelasjonene i noen tilfeller øker.

## 5 Oppsummering

Vi har i dette notatet forsøkt å se på sammenhengen mellom tidsrekker for CR for fem forsikringsområder og fire makroøkonomiske størrelser. Dette har vi gjort ved å studere krysskorrelasjoner og rullerende krysskorrelasjoner, både for de originale tidsrekkene og når vi har fjernet trenden. Fordi tidsseriene fra Gjensidige er korte og har grov tidsoppløsning er det ikke mulig å trekke noen konklusjoner basert på disse studiene.

Vi forsøkte også å gjøre det samme når tidsseriene var forskjøvet i forhold til hverandre i tid. Det var ikke mulig å se av dette om korrelasjonene økte systematisk, dvs. at de makroøkonomiske størrelsene var ledende for Gjensidiges variable.

I appendiks A går vi gjennom metodikk som er aktuell for å studere sykler og korrelasjoner med et bedre datagrunnlag. Gjensidiges tidsserier må være betydelig lengre enn de vi har hatt tilgang til, gjerne dobbelt så lange. Videre ville det være en stor fordel å ha kvartalsvis eller månedlig tidsoppløsning. Med et større datagrunnlag kunne vi gjort studier av samme type som Bjørnland (2000). Vi ville da modellert trend og sesong, fjernet disse komponentene og deretter sett på samvariasjon og sykler i restleddene. Videre ville vi undersøkt om det var sammenheng mellom trendene i de ulike seriene ved å benytte kointegrasjon.

# A Mer metode

I dette kapittelet presenterer vi metodikk som er relevant for en analyse av sykler i Gjensidiges data og sammenhengen mellom Gjensidiges data og makroøkonomiske variable. Metodene er ikke benyttet i dette notatet fordi vi fant at tidsseriene fra Gjensidige var for korte og hadde for grov oppløsning.

## A.1 Test av stasjonaritet

Det fins mange tester for stasjonaritet av en tidsserie. Vi vil fokusere på to; *augmented Dickey-Fuller* (ADF)-testen og *Phillips-Perron* (PP)-testen. For å kunne forstå disse, må vi vite hva en integrert tidsserie av orden  $p$  er, samt begrepet enhetsrot (unit root).

Anta at vi har en generell ARMA( $p,q$ ) tidsserie

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad (\text{A.1})$$

der  $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ , se f.eks. Hamilton (1994). Det er vanlig å innføre en lagoperator  $L$  som er slik at  $L^k y_t = y_{t-k}$  for et generelt heltall  $k$ . Vi kan da skrive (A.1) på formen

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \epsilon_t. \quad (\text{A.2})$$

Den karakteristiske ligningen  $1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$  kan faktoriseres som  $(1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_p z)$ . Dersom alle røttene  $z = 1/\lambda_i$ ,  $i = 1, \dots, p$ , ligger utenfor enhetssirkelen, eksisterer den inverse og vi får at

$$y_t = \mu + \psi(L) \epsilon_t,$$

der  $\mu = c / (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$  og  $\psi(L) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) / (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ . Kravene til stasjonaritet er da oppfylt, og vi sier at tidsserien er integrert av orden 0,  $I(0)$ . Dersom den aktuelle tidsserien i tillegg har en deterministisk trend, vil tidsserien da være trendstasjonær. Videre, hvis  $z = 1$  er en rot av den karakteristiske ligningen med multiplisitet 1, får vi at  $(1 - L)y_t = \mu^* + \psi^*(L)\epsilon_t$ . Vi sier da at tidsserien har en enhetsrot og er integrert av orden 1,  $I(1)$ . Generelt er en tidsserie integrert av orden  $d$ ,  $I(d)$ , hvis  $z = 1$  er en rot med multiplisitet  $d$ . Tallet  $d$  angir hvor mange ganger en er nødt til å differensiere tidsserien før den er stasjonær (eller trendstasjonær).

Dersom en har en ARMA( $p,q$ ) som er integrert av orden  $d$ , har vi en ARIMA( $p,d,q$ ). Said og Dickey (1984) viser at det mulig å tilpasse en ARIMA( $p,1,q$ ) til en autoregresjon, der orden  $r$  er en funksjon av antall observasjoner  $n$ :

$$y_t = DT + \phi y_{t-1} + \sum_{j=1}^r \psi_j \Delta y_{t-j} + \epsilon_t. \quad (\text{A.3})$$

$DT$  er en fellesbetegnelse for alle deterministiske ledd, som f.eks. en konstant eller en trend. Ved å bruke minste kvadraters metode for å tilpasse de observerte verdiene til denne testregresjonen kan en finne en teststatistikk som har en gitt grensefordeling. Det er med andre ord mulig å teste nullhypotesen om at tidsserien er  $I(1)$  mot alternativet at den er  $I(0)$  uten å vite verdiene for  $p$  og  $q$ . Dette er i korte trekk forklaringen til ADF-testen. Det er avgjørende å finne en passende verdi for ordenen  $r$  til den tilpassede serien. En tommelfingerregel er å velge  $r$  slik at feilleddene er serielt ukorrelerte. Dersom  $r$  er for liten, vil korrelerte feilledd gi skjevhet, og dersom  $r$  er for stor, vil styrken til testen bli dårligere. Leseren henvises til f.eks. Said og Dickey (1984), Zivot og Wang (2003) og Dickey og Fuller (1979) for en nærmere forklaring.

Phillips-Perron (PP)-testen skiller seg fra ADF-testen i måten den håndterer seriell korrelasjon i feilleddene, samt tilfeller der feilleddene ikke er uavhengig og identisk fordelte. Testregresjonen for PP-testen er

$$y_t = DT + \phi y_{t-1} + u_t, \quad (\text{A.4})$$

der  $u_t$  er  $I(0)$  og ikke nødvendigvis uavhengig og identisk fordelte. PP-testen korrigerer for seriell korrelasjon og feilledd som ikke er uavhengig og identisk fordelte ved å modifisere teststatistikken direkte. Dette er den store fordel ved å bruke PP-testen framfor ADF-testen. En annen fordel er at brukeren ikke trenger å spesifisere noen orden til testregresjonen.

## A.2 Fjerning av trend

Det kan være interessant å se på samvariasjon utover det at makrovariablene og Gjensidiges variabler har en felles trend. Med trend menes en bestemt utvikling eller mønster som tidsserien varierer rundt. Lineær vekst er et eksempel på en trend, mens faste sesongvariasjoner er en annen type trend. Bjørnland (2000) m.fl. har påpekt at fjerning av trend i noen tilfeller kan introdusere et syklisk mønster i restleddene. Dette fenomenet kalles "falske sykler" og reflekterer at metodikken introduserer et mønster i data utover det som finnes i den originale tidsrekken. Nelson og Kang (1981) har f.eks. vist at dette kan skje dersom man detrender data som faktisk er generert av en random-walk på en uheldig måte.

Trenden kan fjernes ved å benytte avanserte filter- eller spektralteknikker slik som f.eks. Bjørnland (2000). I noen tilfeller gir det mening å fjerne eventuelle lineære trender ved å tilpasse en lineær regresjon, dvs.

$$y_t = a + bt.$$

I andre tilfeller er det nødvendig å modellere trenden for å kunne fjerne den.

### A.3 Fjerning av sesong

Dersom man ønsker å studere samvariasjon mellom makrovariable og Gjensidiges variabler utover den sammenhengen som skyldes felles sesongvariasjon, er det viktig å fjerne sesongvariasjoner både fra de makroøkonomiske størrelsene og fra Gjensidiges variabler. Flere av de makroøkonomiske størrelsene som SSB har finnes i en sesongjustert variant.

Med sesongjustering, menes en dekomponering av tidsrekken i en sesongkomponent og en "rest", dvs.

$$y_t = s_t + x_t,$$

der  $s_t$  er sesongkomponenten og  $x_t$  er restleddet. Sesongkomponenten er i seg selv interessant, men i våre analyser vil fokus være på restleddet.

Det finnes flere metoder for å trekke ut sesongkomponenten. Et alternativ er metoden til Cleveland et al. (1990) som er implementert i funksjonen `st1` i S-Plus. Kort fortalt så baserer denne metoden seg på estimere sesongen ved å benytte en sekvens av lokale regresjonsmodeller.

Et alternativ til Cleveland et al. (1990) er å estimere  $s_t$  ved å tilpasse sinus og cosinus funksjoner med forskjellige perioder.

### A.4 Kointegrasjon

Kointegrasjon måler sammenhengen mellom tidsserier ved å beskrive hvordan trendene i seriene påvirker hverandre. Metoden benytter altså tidsrekkene slik de er, evt. justert for sesong, men uten at trenden er fjernet. Begrepet kointegrasjon omfatter flere klasser av modeller som kan ha både enkle og relativt kompliserte avhengighetsstrukturer. Dette betyr at kointegrasjon og korrelasjon er ulike konsepter selv om begge måler samvariasjon.

Hvis man tar utgangspunkt i at det er de makroøkonomiske variablene som eventuelt driver Gjensidiges størrelser (og ikke motsatt), kan et relativt enkelt rammeverk benyttes for parvis kointegrasjon. Dette rammeverket kalles ofte Engle-Granger metoden (Engle og Granger, 1987).

I denne settingen vil vi si at Gjensidiges variabel  $g_t$  og den makroøkonomiske variabelen  $m_t$  er kointegrert dersom

$$g_t = c + \alpha m_t + \epsilon_t, \tag{A.5}$$

og  $\epsilon_t$  er en stasjonær tidsrekke. Et alternativ er å inkludere en drift over tid i regresjonsmodellen, dvs.

$$g_t = c + \delta t + \alpha m_t + \epsilon_t, \tag{A.6}$$

Man kan også undersøke om  $g_t$  er kointegrert med  $m_{t-k}$  for et gitt "lag"  $k$  eller om  $g_t$  er avhengig av den makroøkonomiske størrelsen på flere tidspunkt enn ett.

I praksis kan vi teste om  $g_t$  og den makroøkonomiske variabelen  $m_t$  er kointegrert ved å benytte følgende fremgangsmåte:

- Dersom variabelen har en sesongavhengighet, må den sesongjusteres.
- Tilpass regresjonen (A.5) eller (A.6) ved å benytte vanlig lineær regresjon.
- Beregn residualene  $\hat{\epsilon}_t$ .
- Sjekk om disse er stasjonære ved å benytte en justert unit root test, se Zivot og Wang (2003, pp. 432–434) og appendiks A.1.



# Figurer

1	Bolig CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	8
2	Landbruk CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	9
3	Yrkesskade CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	10
4	Uførhet CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	11
5	Motorvogn CR. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	12
6	KPI. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	14
7	BNP. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	15
8	Lønn. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	16

9	Arbeidsledighetsprosent. Øverst til venstre er den originale serien, øverst til høyre vises logaritmen til etterfølgende differanser, nederst til venstre er ACF til differansene plottet, og nederst til høyre vises PACF til differansene. . . . .	17
10	Bolig CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Bolig CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	23
11	Landbruk CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Landbruk CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	24
12	Yrkesskade CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Yrkesskade CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	25
13	Uføre CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Uføre CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	26
14	Motor CR mot alle makroøkonomiske variable. Kolonnen til venstre viser Motor CR data plottet mot hver av de makroøkonomiske variablene, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	27
15	Bolig CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Bolig CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	28



16	Landbruk CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Landbruk CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	29
17	Yrkesskade CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Yrkesskade CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	30
18	Uføre CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Uføre CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	31
19	Motor CR mot alle makroøkonomiske variable, der trendene er tatt bort. Kolonnen til venstre viser Motor CR data plottet mot hver av de makroøkonomiske variablene, alle uten trend, mens kolonnen til høyre viser de tilhørende rullerende krysskorrelasjonene når seriene ikke er forskjøvet i forhold til hverandre. . . . .	32



# Tabeller

1	Tabell over skadedata fra Gjensidige . . . . .	13
2	Tabell over makroøkonomiske variable . . . . .	13



# Referanser

Bjørnland, H. C. (2000). Detrending methods and stylized facts of business cycles in Norway – an international comparison. *Empirical Economics*, 25:369–392.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. og Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of official statistics*, 6:3–73.

Dickey, D. A. og Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Engle, R. F. og Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276.

Hamilton, J. D. (1994). *TimeSeriesAnalysis*. Princeton University Press.

Kåresen, K. F. og Laading, J. K. (2000). Statistical Market-adjusted Auto-Regressive Term model; Simulation of Long-Term El-Spot Prices. Technical Report SAMBA/02/00, Norwegian Computing Center.

MathSoft, I. (2000). *S-PLUS 6, Guide to Statistics, Vol 2*. MathSoft, Inc.

Nelson, C. R. og Kang, K. H. (1981). Spurious periodicity in inappropriately detrended time series. *Econometrica*, 49:741–751.

Said, S. E. og Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71:599–607.

Zivot, E. og Wang, J. (2003). *Modeling financial time series with S-PLUS*. Insightful.