6: Spuriøs regression og kointegration

Tidsrækkeanalyse

Kasper Rosenkrands

Spuriøs regression

En spuriøs sammenhæng ses mellem to variable som er korrellerede men ikke kan bruges til at forklare korrellationen.

Et eksempel på en spuriøs sammenhæng er samvariationen mellem antal isvafler der sælges og antal drukneulykker. Sammenhængen er spuriøs da vi ikke kan konkludere at man drukner af at spise en isvaffel, der er nok nærmere en underliggende variabel som påvirker begge disse variable (årstiden for eksempel).

I tidsrækkeanalyse kaldes det for **spuriøs regression** når man kigger på sammenhænge mellem uafhængige ikke-stationære tidsrækker.

Måder at undgå det spuriøse regressions problem

En måde at undgå spuriøs regression er at differense de I(1) tidsrækker man betragter, for på den måde at åbne noget stationært.

Denne tilgang er dog ikke altid tilstrækkelig i visse økonomiske anvendelser hvor information om lang tids sammenhænge vil gå tabt ved at gå fra at betragte for eksempel priser til at betragte afkast.

Derfor bliver man nødt til at finde en måde at analysere sammenhænge mellem ikke stationære tidsrækker uden at differense. Det kan man gøre ved hjælp af **kointegration**.

Kointegration

Indgangene i en vektor $\mathbf{x}_t$ siges at være **kointegrerede** af orden d, b, skrives $\mathbf{x}_t \sim CI(d, b)$, if

1. alle indgange i vektoren $\mathbf{x}_t$ er I(d),
2. en vektor $\boldsymbol{\alpha}(\neq 0)$ eksisterer sådan at
   $z_t = \boldsymbol{\alpha}'\mathbf{x}_t \sim I(d-b),\ \ b > 0,$

vektoren $\boldsymbol{\alpha}$ kaldes for den kointegrerende vektor.

For $d = 1, b = 1$ får vi netop at $z_t \sim$ I(0) (stationær).

Test for kointegration

Jeg vil nu præsentere Engle-Grangers 2-step metode til at bestemme om (for simpelthedens skyld) 2 I(1) tidsrække kointegrerer.

Step 1

1. Tjek at begge tidsrækker er I(1).
2. Estimer en regression med en af tidsrækker som respons og den anden som forklarende variabel

$$x_{1t} = \mu + \beta_2 x_{2t} + w_t,$$

   gem resiudalerne i $\hat{w}_t$.

3. Test om $\hat{w}_t$ er I(1) eller I(0).

Hvis $\hat{w}_t$ er I(1)

Så betyder det at de to tidsrækker **ikke** kointegrerer og man kan i så fald estimere en model der kun indeholde 1. differenser af de oprindelige tidsrækker, med andre ord en model der forklarer **kortsigts-sammenhænge**.

Hvis $\hat{w}_t$ er I(0)

Så betyder det at de to tidsrækker **kointegrerer** og vi kan da gå til trin 2 for at estimere en **fejlkorrektionsmodel (ECM)**.

Step 2

1. Brug step 1 residualerne fra forrige periode som en variabel i fejlkorrektionsmodellen:

$$\nabla x_{1t} = \psi_0 + \psi_1 \nabla x_{1,t-1} + \underbrace{\psi_2 \nabla x_{2,t-1}}_{\text{kortsigts}} + \underbrace{\lambda \hat{w}_{t-1}}_{\text{langsigts}} + v_t,$$

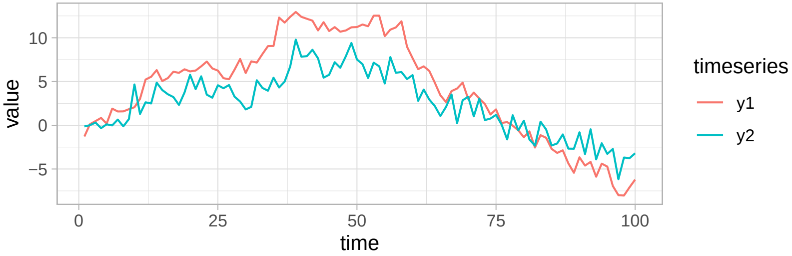
   hvor $\hat{w}_{t-1} = x_{1,t-1} - \hat{\mu} - \hat{\beta}_2 x_{2,t-1}$.

2. Værdien af λ bestemmer fejlkorrektions-hastigheden, den vil altid være negativ da vi ellers vil divergere fra langsigts ligevægt.

Kointegration eksempel

Her har jeg simuleret to tidsrækker,

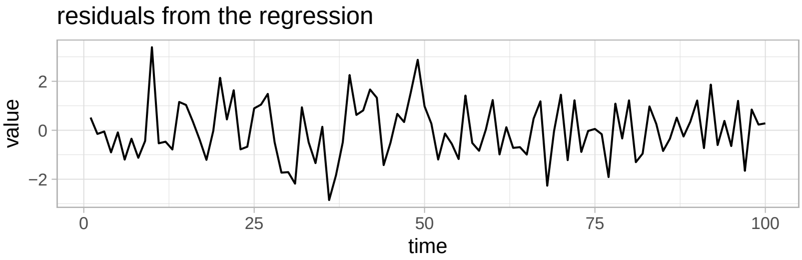
- y_1 er en random walk
- y_2 er givet ved
$$y_{2t} = .6y_{1t} + w_t,$$
hvor w_t er hvid støj.



Test for kointegration eksempel

Vi ser direkte at de to tidsrækker er I(1), derfor går vi videre til at estimere en regression med y_2 som respons og y_1 som forklarende variabel.

```r
reg <- lm(y2 ~ y1)
```



Test for enhedsrod

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8632 -0.8014 -0.0485  0.9145  3.4024
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.969222   0.141777  -6.836  7.5e-10 ***
## z.diff.lag  0.005222   0.101994   0.051    0.959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.164 on 96 degrees of freedom
## Multiple R-squared:  0.482,  Adjusted R-squared:  0.4712
## F-statistic: 44.66 on 2 and 96 DF,  p-value: 1.941e-14
##
##
## Value of test-statistic is: -6.8363
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

Engle-Granger Table

[Engle - Granger Table]

Antal variable i ligningen	Sample size	Kritisk værdi 1%	5%	10%
2	50	-4.32	-3.67	-3.28
	100	-4.07	-3.37	-3.03
	200	-4.00	-3.37	-3.02
3	50	-4.84	-4.11	-3.73
	100	-4.45	-3.93	-3.59
	200	-4.35	-3.78	-3.47
4	50	-4.94	-4.35	-4.02
	100	-4.75	-4.22	-3.89
	200	-4.70	-4.18	-3.89
5	50	-5.41	-4.76	-4.42
	100	-5.18	-4.58	-4.26
	200	-5.02	-4.48	-4.18

Kilde:Engle & Yoo (1987): Forecasting and testing in cointegrated systems, *Journal of Econometrics* 35, side 143-159.

Estimation af ECM model

Fejlkorrektionsmodellen er givet ved

$$\nabla x_{1t} = \psi_0 + \psi_1 \nabla x_{1,t-1} + \psi_2 \nabla x_{2,t-1} + \lambda \hat{w}_{t-1} + v_t,$$

hvor $\hat{w}_{t-1} = x_{1,t-1} - \hat{\mu} - \hat{\beta}_2 x_{2,t-1}$.

```
##
## Call:
## lm(formula = dy2 ~ dy1.1 + dy2.1 + error.lagged, data = diff.dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7161 -0.8896 -0.0201  0.6889  3.6778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.05316    0.12747  -0.417   0.6776
## dy1.1         0.33576    0.13381   2.509   0.0138 *
## dy2.1        -0.95611    0.11087  -8.624 1.54e-13 ***
## error.lagged -0.92878    0.15349  -6.051 2.92e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.26 on 94 degrees of freedom
## Multiple R-squared:  0.4453, Adjusted R-squared:  0.4276
## F-statistic: 25.15 on 3 and 94 DF,  p-value: 4.931e-12
```