

Introduktion til Statistiske Modeller
for Finansielle Tidsserier

Forelæsningsnoter til Finansiell Økonometri

Jesper Lund

mail@jesperlund.com

<http://www.jesperlund.com>

14. marts 2006

1 Indledning

Formålet med denne forelæsningsnote er at beskrive nogle statistiske tidsseriemodeller, som ofte anvendes indenfor finansiel økonometri. Det er primært modellernes matematiske struktur, som beskrives her. For så vidt angår eksempler på anvendelser henvises til CLM [Campbell *et al* (1997)] og forelæsningerne i faget.

GARCH modeller behandles ikke her, idet der senere kommer en særskilt forelæsningsnote om dette emne. For en uddybende fremstilling af emnet tidsseriemodeller henvises til lærebøger som Hamilton (1994) og Harvey (1993).¹

2 Autokovarianser og autokorrelationer

En tidsserie $\{y_t\}$ er en sekvens (følge) af stokastiske variable, som observeres på forskellige tidspunkter t . Den statistiske model, som beskriver udviklingen i denne tidsserie kaldes en stokastisk proces eller tidsseriemodel.² Den grundlæggende egenskab ved en stokastisk proces er afhængigheden mellem observationerne på forskellige tidspunkter. Ofte er vi kun interesseret i den lineære afhængighed mellem y_t 'erne. Til dette formål bruges kovariansen mellem y_t og y_{t-k} , som er defineret ved

$$\text{Cov}(y_t, y_{t-k}) = E \left[(y_t - E(y_t)) (y_{t-k} - E(y_{t-k})) \right]. \quad (1)$$

For $k = 0$ er dette variansen på y_t . Vi vil primært betragte stationære stokastiske processer, hvor y_t og y_{t-k} har den samme marginale fordeling, og hvor afhængigheden mellem de to stokastiske variable kun afhænger af afstanden mellem de respektive tidspunkter, dvs. k , der kaldes laglængden.

Når vi antager stationaritet, kan vi introducere begrebet *autokovariansfunktionen*, som er ligning (1) som funktion af laglængden k ,

$$\gamma(k) = E \left[(y_t - \mu) (y_{t-k} - \mu) \right], \quad \text{hvor } \mu = E(y_t). \quad (2)$$

I de følgende afsnit vil opstille generelle udtryk for autokovariansfunktionen for de forskellige modeller (stokastiske processer).

Ofte er det mere naturligt at benytte korrelationer i stedet for kovarianser. Derfor introducerer vi *autokorrelationsfunktionen* $\rho(k)$, der er defineret ved:

$$\rho(k) = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var}(y_t)}\sqrt{\text{Var}(y_{t-k})}} = \gamma(k)/\gamma(0), \quad (3)$$

idet y_t og y_{t-k} har samme standardafvigelse, nemlig kvadratroden af variansen $\gamma(0)$. For en given tidsserie (stikprøve), kan autokorrelationsfunktionen estimeres ved hjælp af empiriske varianser og kovarianser for y_t og y_k , f.eks.

$$\hat{\gamma}(k) = \frac{1}{T} \sum_{t=k+1}^T (y_t - \hat{\mu}) (y_{t-k} - \hat{\mu}), \quad \text{hvor } \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t. \quad (4)$$

¹Harvey (1993) og Hamilton (1994) er mine personlige favoritter, men der er naturligvis skrevet mange andre lærebøger om emnet tidsseriemodeller.

²Ordene stokastisk proces og tidsseriemodel bruges i flæng i denne forelæsningsnote.

Man kan argumentere for at der kun skal divideres med $T - k - 1$ for at få et unbiased estimat, men stort set alt statistisk teori for tidsserier er asymptotisk, dvs. at resultaterne gælder formelt kun for $T \rightarrow \infty$.

Den empiriske, eller estimerede, autokorrelationsfunktion er på tilsvarende vis defineret ved:

$$\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0). \quad (5)$$

Generelle udtryk for variansen på estimatorerne (4) og (5) er relativt komplicerede, da samtlige autokorrelationer indgår, se ligningerne (2.4.8) og (2.4.10) i CLM. Hvis alle sande autokorrelationer er 0, er den asymptotiske varians på $\hat{\rho}(k)$ dog meget simpel, nemlig

$$\text{Var}[\hat{\rho}(k)] = \frac{1}{T}, \quad \text{hvis } \rho(j) = 0 \text{ for alle } j \geq 1. \quad (6)$$

Som regel har man kun brug for standardafvigelsen på $\hat{\rho}(k)$ når man skal teste en nulhypotese om ingen afhængighed, dvs. $\rho(k) = 0$ for alle k , og hertil er formlen i ligning (6) tilstrækkelig.

3 Hvid støj

Den simpleste stokastiske proces er selvfølgelig ingen tidsmæssig afhængighed, hvilket ofte kaldes *hvid støj* (engelsk: white noise) i tidsserielitteraturen.³ Hvis $\{\varepsilon_t\}$ er en hvid støj process, er $\rho(k) = 0$ for alle $k > 0$.

Ofte antages det at middelværdien er 0, og at den stokastiske variabel ε_t er normalfordelt med varians $\gamma(0) = \sigma^2$. Med disse yderligere antagelser har ε_t de sædvanlige egenskaber for et fejld i en regressionsmodel.

Normalfordelt hvid støj, eller blot hvid støj med middelværdi 0, er den basale byggeklods i AR(p), MA(q) og ARMA(p, q) modeller, som behandles i de efterfølgende afsnit i denne forelæsningsnote.

4 Betingede fordelinger og likelihood funktioner

Hvis x og y er to stokastiske variable med simultan sandsynlighedsfordeling $p(x, y)$, er den betingede fordeling for y givet x defineret ved

$$p(y|x) = \frac{p(x, y)}{p(x)}. \quad (7)$$

Hvis x og y er stokastisk uafhængige, er $p(y|x) = p(y)$, altså den betingede fordeling er identisk med den marginale fordeling. Det svarer til at den simultane fordeling kan skrives som $p(x, y) = p(y)p(x)$.

³Begrebet "hvid støj" stammer fra ingeniørvidenskaben, hvor store dele af den statistiske teori for tidsserier er udviklet.

I mange situationer inden for tidserieanalyse har man brug for et mere generelt udtryk for afhængighed over tid end korrelation, som måler den lineære afhængighed. Til dette formål bruges den betingede fordeling eller, mere typisk, momenter fra den betingede fordeling.

En *martingale differens* proces (MD proces) er defineret ved følgende egenskab:

$$E(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1) = \int \varepsilon_t p(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1) d\varepsilon_t = 0 \text{ for alle } t, \quad (8)$$

hvor $p(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1)$ er den betingede tæthedsfunktion for ε_t givet alle de tidligere værdier af processen.

En MD proces udelukker ikke at der er stokastisk afhængighed mellem ε_t og ε_{t-k} , men den betingede middelværdi er uafhængig af ε_{t-k} . En MD proces opfylder egenskaberne for en hvid støj proces uden antagelsen om normalitet. Omvendt kan man ikke konkludere, at en hvid støj proces er en MD proces.⁴ Normalfordelt hvid støj vil dog altid være en MD proces, idet uafhængighed og ingen korrelation er det samme i den simultane normalfordeling.

Ved maximum likelihood estimation bruger man ofte følgende omskrivning af den simultane tæthed for alle observationer:

$$\begin{aligned} p(y_1, y_2, \dots, y_T) &= p(y_T | y_{T-1}, \dots, y_2, y_1) p(y_{T-1}, \dots, y_2, y_1) \\ &= p(y_T | y_{T-1}, \dots, y_2, y_1) p(y_{T-1} | y_{T-2}, \dots, y_2, y_1) p(y_{T-2}, \dots, y_2, y_1) \\ &= \left\{ \prod_{t=2}^T p(y_t | Y_{t-1}) \right\} p(y_1), \text{ hvor } Y_t = (y_t, y_{t-1}, \dots, y_2, y_1). \end{aligned} \quad (9)$$

Den sidste linie i ligning (9) fremkommer ved gentagen anvendelse af spaltningen i de to første linier. Det smarte ved omskrivningen i ligning (9) er at en tidsseriemodel generelt er specificeret ved den betingede tæthedsfunktion $p(y_t | Y_{t-1})$. Ud fra den betingede tæthedsfunktion kan man let konstruere den simultane tæthedsfunktion, og dermed likelihoodfunktionen. Tricket i ligning (9) kaldes *prediction error decomposition* i litteraturen.

Forudsigelsesfejlen (prediction error) er defineret som

$$u_t = y_t - E(y_t | Y_{t-1}), \quad (10)$$

og den statistiske model vil typisk specificere en fordeling for u_t og den funktionelle form af den betingede middelværdi for y_t givet de tidligere værdier af processen, dvs. Y_{t-1} med vores notation. Den stokastiske variabel u_t kaldes ofte en *innovation* i tidsserielitteraturen.

Log-likelihood funktionen kan nu skrives som

$$\begin{aligned} \log L(y_1, y_2, \dots, y_T) &= \sum_{t=h+1}^T \log p(y_t | Y_{t-1}) + \log p(y_h, \dots, y_1) \\ &\approx \sum_{t=h+1}^T \log p(y_t | Y_{t-1}) \end{aligned} \quad (11)$$

⁴En GARCH model er et eksempel på en stokastisk proces, som kan være MD uden at være hvid støj, altså hvor der kan være stokastisk afhængighed men lineær uafhængighed (ingen korrelation).

I den første linie af (11) er der valgt en lidt mere generel opskrivning, hvor den simultane tæthed for de første h observationer ikke spaltes i produktet af de betingede tætheder. Logaritmen til tætheden (9) vil således svare til $h = 1$ i (11). Den anden linie i (11) er en approksimation til log-likelihood funktionen, idet leddet $\log p(y_h, \dots, y_1)$ er udeladt. I mange situationer er den approksimative log-likelihood funktion nemmere at håndtere, og hvis antallet af observationer T er tilpas stort, vil der ikke være en nævneværdig forskel mellem at maksimere den eksakte log-likelihood funktion og den approksimative. For eksempel kan en $AR(p)$ estimeres ved hjælp af almindelig OLS, hvis man bruger den approksimative log-likelihood funktion med $h = p$, mens den eksakte log-likelihood funktion er noget mere kompliceret at håndtere.

5 MA(q) processer

En MA(q) proces defineres med udgangspunkt i en hvid støj ε_t , eventuelt en normalfordelt hvid støj af hensyn til muligheden for at opstille log-likelihood funktionen. MA(q) er defineret ved

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \\ &= \mu + \sum_{j=0}^q \theta_j \varepsilon_{t-j}, \quad \text{med } \theta_0 \equiv 1, \end{aligned} \quad (12)$$

og hvor $\varepsilon_t \sim N(0, \sigma^2)$. Middelværdien af y_t er μ , idet $E(\varepsilon_t) = 0$ for alle t . Den simpleste MA(q) proces er MA(1) processen, hvor

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}. \quad (13)$$

Autokovariansfunktionen for en MA(q) proces er givet ved

$$\gamma(k) = \begin{cases} \sigma^2 \sum_{j=k}^q \theta_j \theta_{j-k} & \text{hvis } 0 \leq k \leq q \\ 0 & \text{hvis } k > q. \end{cases} \quad (14)$$

Beviset fremkommer ved at opstille udtrykket

$$\begin{aligned} E[(y_t - \mu)(y_{t-k} - \mu)] &= E\left(\sum_{i=0}^q \theta_i \varepsilon_{t-i} \sum_{j=0}^q \theta_j \varepsilon_{t-k-j}\right) \\ &= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j E(\varepsilon_{t-i} \varepsilon_{t-k-j}) \end{aligned} \quad (15)$$

og udnytte at

$$E(\varepsilon_{t-i} \varepsilon_{t-k-j}) = \begin{cases} \sigma^2 & \text{hvis } i = k - j \\ 0 & \text{ellers.} \end{cases} \quad (16)$$

Hvis $k \leq q$ er der $q - k$ led i summen (15), som er forskellig fra 0. Omvendt er der ingen korrelation mellem y_t og y_{t-k} hvis $k > q$. Den tidsmæssige afhængighed i en MA(q) proces er altså begrænset til de seneste q observationer.

Ved at sætte $k = 0$ i ligning (14) kan variansen bestemmes som

$$\text{Var}(y_t) = \gamma(0) = \sigma^2 \sum_{j=0}^q \theta_j^2 = \sigma^2 \left(1 + \sum_{j=1}^q \theta_j^2 \right). \quad (17)$$

Det indebærer at autokorrelationsfunktionen kan skrives som

$$\rho(k) = \begin{cases} \frac{\sum_{j=k}^q \theta_j \theta_{j-k}}{1 + \sum_{j=1}^q \theta_j^2} & \text{hvis } 0 \leq k \leq q \\ 0 & \text{hvis } k > q. \end{cases} \quad (18)$$

Estimation af parametrene i en $MA(q)$ proces behandles i afsnittet om estimation af ARMA modeller, idet $MA(q)$ modellen er en $ARMA(0, q)$, altså et specialtilfælde.

6 AR(p) processer

AR(p) processen defineres ligeledes med udgangspunkt i en hvid støj ε_t

$$\begin{aligned} y_t &= \mu + \phi_1(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t \\ &= \mu + \sum_{i=1}^p \phi_i(y_{t-i} - \mu) + \varepsilon_t. \end{aligned} \quad (19)$$

Nogle gange ses en alternativ parameterisering, hvor

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t. \quad (20)$$

Det er selvfølgelig den samme model, idet

$$\phi_0 = \mu - \sum_{i=1}^p \phi_i \mu = \mu \left(1 - \sum_{i=1}^p \phi_i \right), \quad (21)$$

hvor $\mu = E(y_t)$, altså den ubetingede middelværdi af y_t .⁵ Et specialtilfælde af AR(p) modellen er AR(1) modellen, hvor

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t. \quad (22)$$

Man kan omskrive en AR(p) model til en $MA(\infty)$, altså en MA model af “uendelig” orden. Det gøres ved successivt at indsætte udtrykket for y_{t-1} i ligning (19). For en AR(1) model med $|\phi| < 1$ giver det

$$\begin{aligned} y_t &= \mu + \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \mu + \phi_1^2(\phi_1 y_{t-3} + \varepsilon_{t-2}) + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \mu + \varepsilon_t + \sum_{i=1}^{\infty} \phi^i \varepsilon_{t-i}. \end{aligned} \quad (23)$$

⁵Dette forudsætter at $\sum_{i=1}^p \phi_i < 1$, hvilket er en af betingelserne for stationaritet. Da vi generelt kun diskuterer stationære stokastiske processer, bortset fra bemærkningerne om I(1) processer i afsnittet om integration og kointegration, vil vi ikke komme nærmere ind på denne problemstilling. Den interesserede læser henvises til Harvey (1993) eller Hamilton (1994) for de præcise betingelser for stationaritet i AR(p) modellen.

Autokovariansfunktionen for en AR(1) model er

$$\begin{aligned}
 \gamma(k) &= E \left[(y_t - \mu)(y_{t-k} - \mu) \right] \\
 &= E \left[\left(\phi^k (y_{t-k} - \mu) + \sum_{i=0}^{k-1} \phi^i \varepsilon_{t-i} \right) (y_{t-k} - \mu) \right] \\
 &= \phi^k E \left[(y_{t-k} - \mu)(y_{t-k} - \mu) \right] \\
 &= \phi^k \gamma(0),
 \end{aligned} \tag{24}$$

idet y_{t-k} er ukorreleret med $\varepsilon_{t-k+1}, \dots, \varepsilon_t$. Autokorrelationsfunktionen for AR(1) modellen er derfor

$$\rho(k) = \gamma(k)/\gamma(0) = \phi^k. \tag{25}$$

Det vil sige at man faktisk slet ikke behøver at kendes den ubetingede varians $\gamma(0)$ for at bestemme autokorrelationsfunktionen. Hvis man får behov for at udlede den ubetingede varians, gøres det nemmest via MA(∞) omskrivningen af AR(1) modellen:

$$\gamma(0) = \sigma^2 + \sum_{i=1}^{\infty} (\phi^i)^2 \sigma^2 = \sigma^2 \sum_{i=0}^{\infty} (\phi^2)^i = \frac{\sigma^2}{1 - \phi^2}, \quad \text{hvis } |\phi| < 1. \tag{26}$$

En alternativ metode til at eftervise dette resultat er at beregne variansen på begge sider af lighedstegnet i ligning (22)

$$\text{Var}(y_t) = \phi^2 \text{Var}(y_{t-1}) + \sigma^2 \implies \gamma(0) = \phi^2 \gamma(0) + \sigma^2, \tag{27}$$

og løse den sidste ligning mht. $\gamma(0)$. Bemærk at vi har udnyttet stationaritetstagselsen, hvorved $\text{Var}(y_t)$ og $\text{Var}(y_{t-1})$ er identiske. Udledningen i (27) kvalificerer dog ikke helt som et bevis for den ubetingede varians, idet man er nødt til at antage stationaritet, mens betingelsen for stationaritet $|\phi| < 1$ udledes samtidig med udtrykket for $\gamma(0)$ i beviset (26).

Man kan også udlede generelle udtryk for autokorrelationsfunktionen i AR(p) processen, men disse udtryk er relativt komplicerede.⁶ I stedet henvises der til lærebøger om tidsseriemodeller som Harvey (1993) og Hamilton (1994).

7 ARMA(p, q) processer

Det er muligt at kombinere AR og MA specifikationerne, hvilket giver en ARMA(p, q) model. Den generelle specifikation er

$$y_t = \mu + \sum_{i=1}^p \phi_i (y_{t-i} - \mu) + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \tag{28}$$

⁶Det er værd at bemærke, at forskellen mellem en AR(2) og AR(1) model er ganske betydelig for så vidt angår udtrykket for autokorrelationsfunktionen, mens der ikke er nævneværdig forskel på en MA(1) og MA(2) model i den henseende. Når vi kommer til estimationen af modellerne, er en AR(2) model dog ikke mere kompliceret end en AR(1) model. Begge kan estimeres vha. OLS, som svarer til betinget maksimum likelihood, hvis ε_t er normalfordelt.

Den simpleste ARMA model er ARMA(1, 1) modellen:

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}. \quad (29)$$

ARMA(p, q) modellen kan ligesom AR(p) modellen omskrives til en MA(∞) model. Dette vises kun for ARMA(1, 1) modellen, hvor vi får

$$y_t = \mu + \varepsilon_t + \sum_{j=1}^{\infty} (\phi + \theta) \phi^{j-1} \varepsilon_{t-j}, \quad (30)$$

altså en MA(∞) model med $\theta_j = (\phi + \theta) \phi^{j-1}$. Hvis $\phi = -\theta$ er $\theta_j = 0$ for alle $j > 0$, og ARMA(1, 1) modellen er identisk med en hvid støj proces. Dette er et eksempel på det såkaldte *identifikationsproblem*. Hvis den sande proces er en hvid støj, kan man godt specificere en ARMA(1, 1) model, men der vil være uendeligt mange forskellige ARMA(1, 1) modeller, som passer til data, nemlig alle modeller hvor $\phi = -\theta$. Dette lyder måske som en noget teoretisk diskussion, men det er i høj grad også et praktisk problem. Hvis man overparameteriserer sin ARMA(p, q) model, altså angiver en for høj orden af både AR og MA delen, er der stor risiko for at estimationen ikke vil konvergere, fordi en AR og MA komponent "ophæver" hinanden.

Autokorrelationsfunktionen for den generelle ARMA(p, q) model er temmelig kompliceret, og den udelades derfor fra denne fremstilling. I faget Finansiell Økonometri vil vi generelt kun støde på rene AR eller MA processer.

8 Estimation af ARMA modeller

Hvis ε_t er normalfordelt med konstant varians, kan man opstille log-likelihood funktionen vha. metoden i afsnit 4. For den generelle ARMA(p, q) er den betingede middelværdi og varians af y_t givet ved

$$E(y_t | Y_{t-1}) = \mu + \sum_{i=1}^p \phi_i (y_{t-i} - \mu) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (31)$$

$$\text{Var}(y_t | Y_{t-1}) = E[(y_t - E[y_t | Y_{t-1}])^2 | Y_{t-1}] = E[\varepsilon_t^2 | Y_{t-1}] = \sigma^2 \quad (32)$$

hvor $Y_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_1)$. Den approksimative log-likelihood funktion for ARMA(p, q) modellen kan nu skrives som

$$\begin{aligned} \log L &= \sum_{t=p+1}^T \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \frac{(y_t - E[y_t | Y_{t-1}])^2}{\sigma^2} \right\} \\ &= \frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=p+1}^T \varepsilon_t^2 \end{aligned} \quad (33)$$

hvor $\varepsilon_t = y_t - E[y_t | Y_{t-1}]$ er innovationen (fejleddet), jf. definitionen i afsnit 4. Bemærk at summationen starter ved observation $t = p + 1$, og ikke $t = 1$. Det skyldes at vi skal bruges p laggede værdier af y_t som startværdier til den betingede middelværdi. Laggede værdier af innovationen ε_t indgår også i den betingede middelværdier. Her

sættes startværdier til 0, dvs. $\varepsilon_t = 0$ for $t \leq p$. Derefter kan den betingede midelværdi, og dermed log-likelihood funktionen, beregnes rekursivt fra $t = p + 1$ til $t = T$.

Log-likelihood funktionen i (33) er ikke den eksakte log-likelihood funktionen, men medmindre antallet af observationer T er meget lille, er approksimationen uden praktisk betydning.⁷ Den approksimative log-likelihood funktion er specielt nyttig for rene AR(p) modeller, hvor (33) er log-likelihood funktionen for en lineær regressionsmodel med y_{t-1}, \dots, y_{t-p} som forklarende variable. Det vil sige, at OLS kan benyttes til estimation. For MA(q) og ARMA(p, q) modeller er det nødvendigt at anvendes numerisk optimering for at finde maximum likelihood estimererne. Statistikprogrammet SAS har rutiner til dette i SAS/ETS.

Før man estimerer en ARMA(p, q) model, skal man naturligvis angive ordenen af AR og MA delen, altså p og q . Umiddelbart kunne det være fristende at angive en meget høj orden af begge dele, og efterfølgende fjerne de insignifikante komponenter vha. t -tests eller F -tests vedr. flere parametre på samme tid. Ved almindelig regressionsanalyse, hvor man er usikker på specifikationen af modellen, kan denne metode ofte anvendes med succes. Metoden er desværre ikke brugbar for blandede ARMA modeller, idet en MA og AR komponent kan "ophæve" hinanden, som vi så i eksemplet med ARMA(1, 1) ovenfor. Dette problem optræder ikke for rene AR eller MA modeller, så her kan "general-to-simple" metoden i princippet anvendes.

For en korrekt specificeret ARMA model gælder det, at fejlleddet ε_t er hvid støj, dvs. alle autokorrelationer er lig med 0. Det kan man teste via residualerne fra den estimerede model.⁸ Ud fra strukturen i autokorrelationsfunktionen $\hat{\rho}(k)$ kan man visuelt vurdere, om det er AR eller MA delen som skal have en højere orden. Denne metode er udviklet af Box & Jenkins (1976), og deres bog var i mange år standardreferencen for ARMA modeller.⁹ Box-Jenkins metoden til specifikation af ARMA modellen diskuteres også af Mills (1990) og Milhøj (1986).

9 Lidt om integration og kointegration

CLM kommer visse steder ind på *enhedsrødder* (unit roots) og stationære sammenhænge mellem to forskellige tidsserier med enhedsrod, hvilket kaldes *kointegration*. I dette afsnit af forelæsningsnoten gives en meget kortfattet introduktion til disse begreber.

⁷Eksakt maximum likelihood estimation kræver generelt, at man inverterer den fulde $T \times T$ kovariansmatrix for samtlige observationer y_1, y_2, \dots, y_T (PROC ARIMA i SAS/ETS gør dette, når man specificerer maximum likelihood estimation med option METHOD=ML). Det er tidskrævende, hvis antallet af observationer T er stort. Man kan dog undgå at invertere en $T \times T$ matrix ved at anvende Kalman filteret til beregning af den eksakte log-likelihood funktion, hvilket diskuteres i kapitel 13 i Hamilton (1994).

⁸Her skal der dog tages højde for at residualerne har lidt andre statistiske egenskaber end fejlleddet ε_t , idet residualerne er beregnet med estimerer på de ukendte parametre. De nærmere detaljer ligger uden for rammerne af denne forelæsningsnote.

⁹I visse dele af tidsserielitteraturen kaldes ARMA modeller ligefrem Box-Jenkins modeller!

9.1 Test for enhedsrødder¹⁰

I den simple AR(1) model med middelværdi 0

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (34)$$

er maximum likelihood estimatet på ρ givet ved

$$\hat{\rho} = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2} \quad (35)$$

Det kan endvidere vises, at den asymptotiske fordeling for ML estimatoren er¹¹

$$\sqrt{T}(\hat{\rho} - \rho_0) \xrightarrow{L} N(0, 1 - \rho_0^2), \quad \text{hvis } |\rho_0| < 1, \quad (36)$$

hvor ρ_0 er den sande værdi af parameteren ρ (den værdi, som vi prøver at estimere). I praksis vil (36) sige, at man kan teste hypoteser om parameteren ρ ved at udnytte at $\hat{\rho}$ er normalfordelt med variansen $\text{Var}(\hat{\rho}) = (1 - \hat{\rho}^2)/T$. Alternativt kan man lave et almindeligt t -test baseret på den standardafvigelse, som et standard OLS program¹² beregner, dvs.

$$t = \frac{\hat{\rho} - \rho_0}{\hat{\sigma}_\varepsilon^2 / \left(\sum_{t=2}^T y_{t-1}^2\right)}, \quad \text{hvor } \hat{\sigma}_\varepsilon^2 = \frac{1}{T-2} \sum_{t=2}^T (y_t - \hat{\rho} y_{t-1})^2. \quad (37)$$

Dette gælder imidlertid ikke, hvis den sande værdi af parameteren er $\rho_0 = 1$, herunder hvis vi tester en nulhypotese om at $\rho = 0$. I tidsserielitteraturen kaldes $\rho_0 = 1$ tilfældet for en *enhedsrod*, eller *unit root* på engelsk.

Hvis $\rho_0 = 1$ indsættes i ligning (36) får man en asymptotisk varians på 0, hvilket naturligvis ikke giver mening. Det fremgår også af ligning (36), at den asymptotiske fordeling kun gælder hvis $-1 < \rho_0 < 1$. OLS standardafvigelsen på $\hat{\rho}$ i nævneren af (37) kan dog altid beregnes, og standardafvigelsen vil selvfølgelig være strengt større end 0, men den asymptotiske fordeling for t -statistikken (37) vil afvige fra normalfordelingen når $\rho_0 = 1$, som vi skal se i det følgende.

Konsekvensen af dette er at $\hat{\rho}$ har en anden asymptotisk fordeling hvis $\rho_0 = 1$. I kapitel 17 i Hamilton (1994) udledes følgende asymptotiske fordelinger for $\hat{\rho}$ og t -testet hvis $\rho_0 = 1$:

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{(1/2) \{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr} \quad (38)$$

$$t \xrightarrow{L} \frac{(1/2) \{[W(1)]^2 - 1\}}{\left\{ \int_0^1 [W(r)]^2 dr \right\}^{1/2}} \quad (39)$$

¹⁰ Dette afsnit kan med fordel læses sammen med afsnit 2.7 i CLM.

¹¹ Formelt er det nødvendigt at udlede den asymptotiske fordeling for $\sqrt{T}(\hat{\rho} - \rho_0)$, idet fordelingen for $\hat{\rho}$ kollapser til en enkelt værdi, nemlig ρ_0 (ellers ville estimatoren ikke være konsistent). Skaleringen med \sqrt{T} er standardmetoden til at udlede den asymptotiske fordeling for en estimator.

¹² For eksempel PROC REG i SAS eller Tools:Data Analysis:Regression i Excel.

De nye asymptotiske fordelinger er funktionaler af en Brownian motion $W(r)$, se kapitel 17 i Hamilton (1994) for en uddybende forklaring af dette begreb. Her skal vi blot bemærke at fordelingerne er mere venstreskæve end normalfordelingen, hvilket betyder at de negative fraktilværdier er lavere end de tilsvarende værdier fra normalfordelingen. For eksempel er 5% fraktilværdien af (39) fordelingen for et ensidet t -test $-1,95$ mod $-1,645$ i normalfordelingen. Populært sagt skal der altså “mere til” for at afvise en nulhypotese pga. den nye fordeling. De specielle statistiske tests baseret på fordelingerne (38) og (39) kaldes ofte for Dickey-Fuller testet i “unit root” litteraturen.

9.2 Kointegration

Testet for $\rho = 1$ i afsnit 9.1 er i virkeligheden et test mellem to vidt forskellige fordelinger, nemlig ikke-stationaritet versus stationaritet. Under nulhypotesen $\rho = 1$ følger y_t random-walk processen

$$y_t = y_{t-1} + \varepsilon_t, \quad (40)$$

som indebærer at der ikke er en stationær fordeling for y_t . Det betyder igen, at stikprøvemomenter som

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t \quad \text{og} \quad \widehat{\text{Var}}(y) = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2 \quad (41)$$

ikke vil konvergere til konstante værdier når $T \rightarrow \infty$ (variansen $\widehat{\text{Var}}(y)$ vil i teorien vokse eksplosivt). Under alternativhypotesen $\rho < 1$ følger y_t derimod en stationær stokastisk process, og den ubetingede middelværdi og varians på y_t eksisterer.

Figur 1 viser grafisk forskellen på en random-walk proces og en stationær autoregression. Selvom det er samme statistiske AR(1) model, blot med hhv. $\rho = 0,97$ og $\rho = 1$ er der meget stor forskel på de to tidsserier. Random-walk processen (den røde tidsserie) har meget store udsving, mens den stationære proces (den blå tidsserie) hele tiden trækkes mod den ubetingede middelværdi 0. En ikke-stationær tidsserie kaldes ofte $I(1)$, mens en stationær kaldes $I(0)$. Betegnelsen $I(1)$ henviser til at der skal tages differenser een gang før processen er stationær, idet førstedifferensen $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$ er en stationær process.

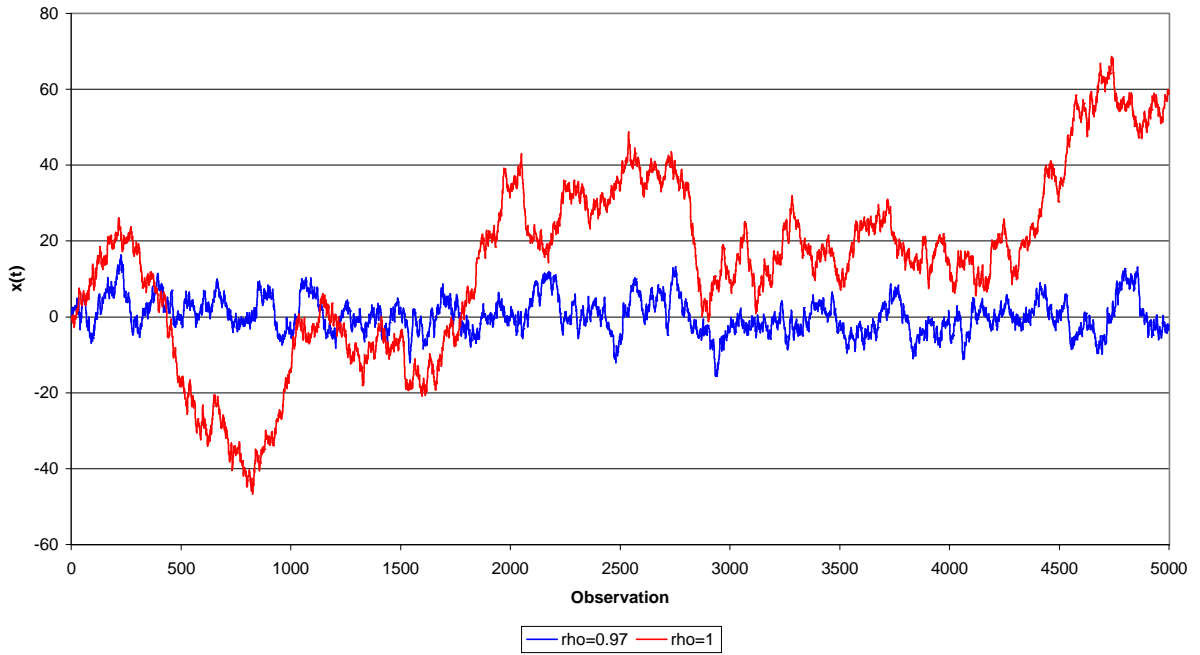
Kointegration er en speciel relation mellem to ikke-stationære tidsserier. Hvis x_t og y_t begge er ikke-stationære, altså $I(1)$, men der eksisterer en lineær relation mellem y_t og x_t som er stationær, altså $I(0)$, siger man at y_t og x_t kointegrerer. Det formelle krav er altså, at der findes en værdi af β , således at

$$e_t = y_t - \beta x_t \sim I(0) \quad (42)$$

selv om y_t og x_t hver for sig er $I(1)$ processer. Ligesom der findes statistiske tests for $I(1)$ mod $I(0)$, findes der statistiske tests for kointegration. Dem skal vi ikke gennemgå her, men blot henvise til eksempelvis Hamilton (1994), hvor dette emne behandles.

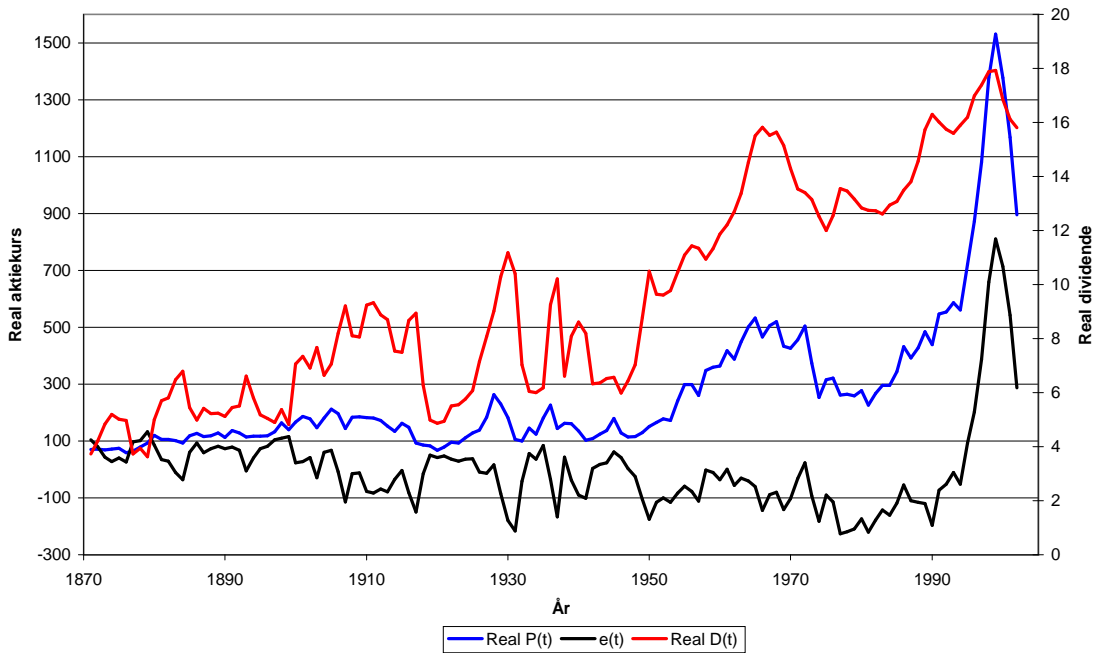
Nogle gange er der økonomisk-teoretiske argumenter for at to tidsserier skal kointegrere. Det gælder f.eks. aktiekurser og dividender, jf. teorien om present-value modeller i CLM kapitel 7. Figur 2 viser den reale aktiekurs og den reale dividende for

Figur 1: Simulation af $x(t)=\rho \cdot x(t-1) + \epsilon(t)$



Figur 1: Forskel på random-walk og stationær AR(1) proces

US reale aktiekurser og dividender



Figur 2: Eksempel på kointegration: reale aktiekurser og dividender i USA

SP-500 aktieindekset i USA. Både aktiekurser og dividender ser umiddelbart ud til at være $I(1)$ processer.¹³ Den tredje tidserie i Figur 2 er residualen e_t fra en OLS regression af aktiekursen på dividendeserien. Denne tidsserie (residualerne) ser umiddelbart ud til at være stationær, hvilket svarer til at der er kointegration mellem aktiekurser og dividender.

10 VAR modeller

VAR betyder Vector AutoRegression, og det er en multivariat generalisering af den univariate AR model. I stedet for en enkelt stokastisk variabel y_t , opstiller vi en tidsseriemodel for en stokastisk vektor \mathbf{x}_t .

Lad \mathbf{x}_t være en n -dimensional stokastisk vektor, altså en $n \times 1$ vektor af stokastiske variable. Vi antager at \mathbf{x}_t kan beskrives ved følgende stokastiske proces

$$\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{A}(\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t, \quad (43)$$

hvor \mathbf{A} er en $n \times n$ matrix af VAR koefficienter (parametre), og $\boldsymbol{\mu} = E(\mathbf{x}_t)$ er den ubetingede middelværdi (middelvektor) af \mathbf{x}_t . Hvis $n = 2$, kan VAR modellen også skrives som

$$x_{1t} = \mu_1 + a_{11}(x_{1,t-1} - \mu_1) + a_{12}(x_{2,t-1} - \mu_2) + \varepsilon_{1t} \quad (44)$$

$$x_{2t} = \mu_2 + a_{21}(x_{1,t-1} - \mu_1) + a_{22}(x_{2,t-1} - \mu_2) + \varepsilon_{2t}, \quad (45)$$

men matrix formen (43) er naturligvis mere generel. Fejlleddet $\boldsymbol{\varepsilon}_t$ i ligning (43) antages at opfylde følgende betingelser:

$$E[\boldsymbol{\varepsilon}_t | \mathbf{X}_{t-1}] = \mathbf{0} \quad (46)$$

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t+1}' | \mathbf{X}_{t-1}] = \boldsymbol{\Omega}, \quad (47)$$

hvor $\mathbf{X}_{t-1} = (\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_1)$ er informationsmængden i den betingede forventning. For at forenkle notationen vil vi nogle gange skrive $E_{t-1}[\cdot]$, der skal fortolkes som

$$E_{t-1}[\cdot | \mathbf{X}_{t-1}]. \quad (48)$$

Ovenstående er en VAR(1) model. Der findes naturligvis også en generel VAR(p) model, der er specificeret som

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^p \mathbf{A}_i (\mathbf{x}_{t-i} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t. \quad (49)$$

I afsnit 10.2 nedenfor skal vi se på forecastfunktionen i en VAR model. Denne funktion er væsentligt nemmere at udlede for en VAR(1) end for en generel VAR(p) model.

¹³Et statistisk test kan i hvert fald ikke afvise at begge processer er $I(1)$.

Derfor bruger man ofte følgende trick, hvorved en VAR(p) kan omskrives til en VAR(1) model på såkaldt *companion* form:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \cdots \\ \mathbf{x}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \cdots \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_{t-1} - \boldsymbol{\mu} \\ \mathbf{x}_{t-2} - \boldsymbol{\mu} \\ \cdots \\ \mathbf{x}_{t-p} - \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \cdots \\ \mathbf{0} \end{bmatrix}. \quad (50)$$

For VAR(2) modellen specialiseres dette til

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} - \boldsymbol{\mu} \\ \mathbf{x}_{t-2} - \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \end{bmatrix}. \quad (51)$$

En VAR(p) model kan således opfattes som en VAR(1) model for den augmentede stokastiske vector

$$\mathbf{Z}_t \equiv \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \cdots \\ \mathbf{x}_{t-p+1} \end{bmatrix} \quad \text{med} \quad \mathbf{Z}_t = \boldsymbol{\mu}^* + \mathbf{A}^* \mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t^* \quad (52)$$

og hvor definitionerne af $np \times 1$ vektorerne $\boldsymbol{\mu}^*$ og $\boldsymbol{\varepsilon}_t^*$, samt $np \times np$ matricen \mathbf{A}^* følger af opskrivningen i ligning (50).

10.1 Estimation af VAR modeller

Hvis $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, kan den eksakte log-likelihood funktion for VAR(p) modellen skrives som

$$\begin{aligned} \log L &= \sum_{t=p+1}^T \left\{ -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \boldsymbol{\varepsilon}_t' \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right\} \\ &\quad - \frac{np}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}_p| - \frac{1}{2} (\mathbf{X}_p - \boldsymbol{\mu}_p)' \mathbf{V}_p^{-1} (\mathbf{X}_p - \boldsymbol{\mu}_p), \end{aligned} \quad (53)$$

hvor \mathbf{X}_p er $np \times 1$ vektoren bestående af de første p observationer af vektoren x_t , og hvor $\boldsymbol{\mu}_p$ og \mathbf{V}_p er hhv. den ubetingede middelvektor og kovariansmatrix for den stokastiske vektor \mathbf{X}_p . Opskrivningen af den eksakte log-likelihood funktion i (53) fremkommer ved at anvende prediction error decomposition metoden, jf. afsnit 4, på den multivariate betingede tæthedsfunktion for \mathbf{x}_t .

Det er meget nemmere at håndtere den approksimative log-likelihood funktion hvor bidraget fra de første p observationer, altså den anden linie i ligning (53) ignoreres. Den approksimative log-likelihood funktion, der også kan fortolkes som den betingede log-likelihood funktion givet de første p observationer på \mathbf{x}_t , kan skrives som:

$$\log L \approx -\frac{(T-p)n}{2} \log(2\pi) - \frac{T-p}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \sum_{t=p+1}^T \boldsymbol{\varepsilon}_t' \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t. \quad (54)$$

Hvis man anvender reparameteriseringen $\mathbf{c} = \boldsymbol{\mu} (\mathbf{I} - \sum_{i=1}^p \mathbf{A}_i)$, svarer dette til log-likelihood funktionen for den multivariate regressionsmodel

$$\mathbf{x}_t = \mathbf{c} + \sum_{i=1}^p \mathbf{A}_i \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, \quad \text{for } t = p + 1, \dots, T. \quad (55)$$

Da der er de samme forklarende variable i alle n regressionsligninger, kan dette system estimeres ligning for ligning ved almindelig OLS uden efficienstab (dvs. OLS ligning-for-ligning er ML estimation).

10.2 Forecast funktionen i en VAR model

I forbindelse med emnet present-value modeller og forventningshypoteseteorien for rentestrukturen skal vi bruge en VAR model til at forudsige fremtidige værdier. Det kaldes forecast funktionen i VAR modellen, og det bedste forecast er den betingede forventning givet den information, som er tilgængelig på forecasttidspunktet.

Den betingede forventning af \mathbf{x}_{t+1} i en VAR(1) model er

$$E_t[\mathbf{x}_{t+1}] = \boldsymbol{\mu} + \mathbf{A}(\mathbf{x}_t - \boldsymbol{\mu}), \quad (56)$$

da $E_t[\boldsymbol{\varepsilon}_{t+1}] = \mathbf{0}$. Den betingede forventning af \mathbf{x}_{t+2} på tidspunkt t kan bestemmes ved at anvende dette princip to gange

$$\begin{aligned} E_t[\mathbf{x}_{t+2}] &= E_t[E_{t+1}[\mathbf{x}_{t+2}]] \\ &= E_t[\boldsymbol{\mu} + \mathbf{A}(\mathbf{x}_{t+1} - \boldsymbol{\mu})] \\ &= \boldsymbol{\mu} + \mathbf{A}\mathbf{A}(\mathbf{x}_{t+1} - \boldsymbol{\mu}), \end{aligned} \quad (57)$$

hvor den første lighed følger af loven om itererede forventninger.¹⁴

Generelt kan det vises at

$$E_t[\mathbf{x}_{t+k}] = \boldsymbol{\mu} + \mathbf{A}^k(\mathbf{x}_t - \boldsymbol{\mu}) \quad (59)$$

hvor \mathbf{A}^k skal læses som

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A}, \quad \mathbf{A}^3 = \mathbf{A}\mathbf{A}\mathbf{A}, \quad \mathbf{A}^4 = \mathbf{A}\mathbf{A}\mathbf{A}\mathbf{A}, \quad \text{etc}, \quad (60)$$

altså matricen \mathbf{A} multipliceret med sig selv $k - 1$ gange. Sagt med andre ord: \mathbf{A}^k er matrix generaliseringen af potensopløftning for skalarer.

¹⁴Hvis \mathcal{F} og \mathcal{G} er informationsæt med relationen $\mathcal{G} \subseteq \mathcal{F}$, gælder det generelt at

$$E[z|\mathcal{G}] = E[E[z|\mathcal{F}]|\mathcal{G}], \quad (58)$$

dvs. at man først kan tage forventningen af den stokastisk variabel z betinget af det store informationsæt \mathcal{F} , og dernæst forventningen af forventningen givet det lille informationsæt \mathcal{G} . Slutresultatet svarer til forventningen af den oprindelige stokastiske variabel givet det lille informationsæt. Dette resultat kaldes *loven om itererede forventninger* (law of iterated expectations).

Litteratur

Box, G.E.P. & G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, revised edition, Holden-Day, San Francisco, CA.

Campbell, J.Y., A.W. Lo & A.C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton NJ.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton NJ.

Harvey, A.C. (1993), *Time Series Models, 2nd edition*, Prentice-Hall, UK.

Milhøj, A. (1986), *Tidsrækkeanalyse for Økonomer*, Akademisk Forlag, København.

Mills, T.C. (1990), *Time Series Techniques for Economists*, Cambridge University Press, Cambridge UK.