Tidsrækkeanalyse – Slides 8

Ege Rubak
(baseret på Slides af Esben Høg)

Institut for Matematiske Fag
Aalborg Universitet

Trend stationaritet

- Mange økonomiske variable er udmærket repræsenteret ved autoregressioner af formen

$$\begin{aligned} y_t &= x_t + \alpha + \beta t, \\ \phi(\mathrm{B}) x_t &= w_t. \end{aligned}$$

- Hvis operatoren (polynomiet) $\phi(\mathrm{B})$ udelukkende indeholder rødder uden for enhedscirklen (altså med længde større end 1), så er variablen y_t en lineær trend plus en stabil (stationær) proces.
- "Shocks" behøver ikke at have persistente (vedvarende) effekter, y_t vil altid vende tilbage til den grundlæggende trend $\alpha + \beta t$.
- Vi siger så, at y_t er *trend* stationær.

Første-ordens integreret proces

- Hvis $\phi(\mathrm{B})$ i repræsentationen

$$y_t = x_t + \alpha + \beta t, \quad \phi(\mathrm{B})x_t = w_t,$$

  har $p - 1$ rødder uden for enhedscirklen men én rod lig med 1, så er den detrendede proces x_t stadigvæk ikke-stationær.
- Men $\nabla x_t = x_t - x_{t-1}$ er stationær (stabil). Og ∇y_t er også stationær.
- y_t siges at være differens-stationær eller første ordens integreret eller I(1)
- "Shocks" har en persistent effekt.

Det basale problem at teste for unit root

- Det at teste en nul-hypotese, at der er en enhedsrod i AR polynomiet er tilsyneladende simpelt, idet det tilsyneladende svarer til en simpel restriktion i en (standard) statistisk model.
- Imidlertid viser det sig i det følgende, at dette strider mod standard regularitetsbetingelser, idet LR-teststørrelsen ikke er asymptotisk χ^2 fordelt.
- To typer af redskaber vil vi bruge:
  1. Sædvanlig visuel inspektion, som i standard tidsrækkeanalyse: Hvis y_t er stationær, så er korrelogrammet af ∇y_t mere kompliceret ("over-differensning"). Hvis y_t er I(1), så er korrelogrammet af ∇y_t simplere end det af y_t.
  2. Formelle statistiske tests, primært stammende fra Dickey & Fuller (1979), findes.
  3. Til det formål at beskrive disse, skal vi have indført en Brownsk bevægelse (ganske kort).

Brownske bevægelser

- Indtil videre har standard asymptotisk teori involveret Gaussiske eller χ^2 fordelinger.
- Denne standardteori dækker mange interessante og praktisk forekommende problemstillinger.
- Men, der er altså nogle tilfælde hvor denne standardteori ikke gælder.
- Eksempler er inferens i tidsrækkeanalyse med integrerede (eller *unit root*) tidsrækker.
- Man har så udviklet en såkaldt ikke-standard asymptotisk teori der involverer kontinuert-tids processer, som kaldes Brownske bevægelser.

Brownsk bevægelse

Definition

En *Brownsk bevægelse* eller en *Wiener proces* er en stokastisk proces $W(t)$ defineret på intervallet $[0,1]$, (altså i kontinuert tid), og som tilfredsstiller

1. $W(0) = 0$.
2. For alle $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \ldots \leq t_k$, så er de stokastiske variable $W(t_1) - W(s_1), \ldots, W(t_k) - W(s_k)$ uafhængige.
3. For alle $s < t$ så er den stokastiske variabel $W(t) - W(s)$ normalfordelt med middelværdi 0 og varians $(t - s)\sigma^2$.
4. Stierne (eng: *trajectories*) er kontinuerte.

Sammenhængen mellem Brownske bevægelser og Random walks

- Betragt en random walk

$$v_t = v_{t-1} + w_t, \text{ for } t = 1, \dots, n,$$

hvor $v_0 = 0$, og w_t er en Gaussisk hvid støj.

- Husk at denne model medfører, at

$$v_t = \sum_{j=1}^{t} w_j.$$

Brownske bevægelser og Random walks (fortsat)

- Til dette formål er det mere bekvemt at arbejde med $\frac{1}{\sqrt{n}} v_t$,

$$n^{-1/2} v_t = n^{-1/2} \sum_{j=1}^{t} w_j,$$

og w_t *i.i.d.* N(0,1) $\Rightarrow n^{-1/2} v_t \sim N(0, t/n)$. Det vil sige

- $n^{-1/2} v_0 = 0$.
- For alle tidspunkter $0 \leq t_1 \leq t_2 \ldots \leq t_k \leq n$ så er forskellene $\{n^{-1/2} v_{t_i} - n^{-1/2} v_{t_{i-1}}, \, i = 2, \ldots, k\}$ indbyrdes uafhængige stokastiske variable med $\{n^{-1/2} v_{t_i} - n^{-1/2} v_{t_{i-1}}\} \sim N(0, t_i/n - t_{i-1}/n)$.
- Så en random walk "efterligner" de første to egenskaber ved en Brownsk bevægelse.

Random Walk

Men

- en random walk er en diskret tids proces defineret på $t = 1, \ldots, n$.
- en Brownsk bevægelse er en kontinuert tids proces defineret på $[0, 1]$.

Derfor, så

- "re-indekserer" man en random walk,
- og indfører en kontinuert tids stokastisk variabel, der indeholder den samme information som en random walk.

Re-indeksering af Random Walk

- I stedet for at indeksere observationerne ved $t = 0, 1, \ldots, n$, så kan vi indeksere dem ved $\tilde{t} = \frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}$.
- Vi kan så skrive $n^{-1/2} v_t$ som en funktion af $\tilde{t}$

$$v_n(\tilde{t}) = n^{-1/2} \sum_{j=1}^{n\tilde{t}} w_j.$$

- Bemærk: re-indekseringen ændrer ikke de statistiske egenskaber:
  - $v_n(0) = 0$.
  - For alle tidspunkter $0 \leq \tilde{t}_1 \leq \tilde{t}_2 \ldots \leq \tilde{t}_k \leq 1$ så er forskellene $\{v_n(\tilde{t}_i) - v_n(\tilde{t}_{i-1}),\ i = 2, \ldots, k\}$ indbyrdes uafhængige stokastiske variable med $\{v_n(\tilde{t}_i) - v_n(\tilde{t}_{i-1})\} \sim N(0, \tilde{t}_i - \tilde{t}_{i-1})$.

Kontinuert-tids funktion

- $\tilde{t}$ definerer et "gitter" (eng: *grid*) af værdier i $[0, 1]$.
- Når $n \to \infty$, så bliver gitteret finere og finere.
- Men $v_n(\tilde{t})$ definerer stadigvæk en funktion i diskret tid.
- Betragt i stedet

$$S_n(t) = n^{-1/2} \sum_{j=1}^{[nt]} w_j,$$

  hvor $[\cdot]$ betegner heltalsdelen.
- $S_n(t)$ er en trappefunktion og altså ikke kontinuert i t.
- Men som $n \to \infty$ så forsvinder disse trin mere og mere, og stien (*sample path*) ligner mere og mere en kontinuert proces (og der er matematisk tekniske måder at gøre stien kontinuert på, som vi ikke skal komme ind på her).

Kort om notation og integraler af Brownske bevægelser

- Nogle integraler over simple funktioner af Brownske bevægelser følger standardfordelinger.
- F.eks. kan det vises, at $\int_0^1 W(t)\mathrm{d}W(t) \sim 0.5\sigma^2\left(\chi^2(1)-1\right)$.
- Andre integraler gør ikke.
- F.eks. gælder der, at $\int_0^1 W^2(t)\mathrm{d}t$ ikke har en standard fordeling. Det kan ikke skrives "pænere".
- Derfor skriver man som regel noget i stil med (hvis det nu er tilfældet): "$\cdots \sim \int_0^1 W(t)\mathrm{d}W(t)$" eller "$\cdots \sim \int_0^1 W^2(t)\mathrm{d}t$".

Dickey-Fuller (DF-0) testen

- Vi betragter først et simpelt tilfælde. Antag at $x_t = \phi x_{t-1} + w_t$ for $t \in \mathbb{N}$ (Nullet i overskriften står så for, at der ikke er konstantled på denne ligning). Vi betragter så nulhypotesen at det er random walk, og alternativet at det er en stationær proces:

$$\mathrm{H}_0: \quad \phi = 1, \qquad \mathrm{H}_A: \quad -1 < \phi < 1.$$

- Vi kan også skrive modellen som $\nabla x_t = \gamma x_{t-1} + w_t$, hvor $\gamma = \phi - 1$, hvorfor hypoteserne svarer til

$$\mathrm{H}_0: \quad \gamma = 0, \qquad \mathrm{H}_A: \quad -2 < \gamma < 0.$$

- Dickey-Fuller testen foreslået af Dickey og Fuller er simpelthen t-teststørrelsen i regressionen af ∇x_t på x_{t-1}.

Fordelingen af DF-0 testen er ikke-standard

- Man ville umiddelbart mene, at t-teststørrelsen for ϕ i en regression

$$\nabla x_t = \gamma x_{t-1} + w_t$$

er t-fordelt for $\gamma = 0$, eller i det mindste asymptotisk $N(0, 1)$.

- Dette gælder imidlertid ikke, idet OLS koefficienten (hvis altså den sande $\gamma = 0$)

$$\hat{\gamma} = \frac{\sum_{t=2}^{n} \nabla x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}$$

ikke opfører sig som $\frac{1}{\sqrt{n}} N(0, \sigma^2)$ for $n \to \infty$.

Fordelingen af DF-0 testen er ikke-standard (fortsat)

- Tælleren konvergerer i fordeling asymptotisk på følgende måde, idet der gælder, at

$$n^{-1} \sum_{t=2}^{n} \nabla x_t x_{t-1} \xrightarrow{F} \sigma^2 \int_0^1 W(r) \mathrm{d} W(r),$$

hvor "$\xrightarrow{F}$" betyder konvergens i fordeling, σ^2 er variansen af ∇x_t, og $W(r)$ angiver en Brownsk bevægelse.

- Nævneren konvergerer i fordeling asymptotisk på følgende måde

$$n^{-2} \sum_{t=2}^{n} x_{t-1}^2 \xrightarrow{F} \sigma^2 \int_0^1 W(r)^2 \mathrm{d} r.$$

- Det vil sige brøken ovenfor konvergerer i fordeling hurtigt, men grænsefordelingen er ikke-standard

$$n\hat{\gamma} \xrightarrow{F} \frac{\int_0^1 W(r) \mathrm{d} W(r)}{\int_0^1 W(r)^2 \mathrm{d} r}.$$

Koefficienter og t-værdier

- Oprindeligt blev en test baseret på $\hat{\gamma}$ foreslået af Dickey & Fuller (1979), sammen med testen baseret på $t-$ teststørrelsen (som altså ikke er t-fordelt).
- Senere er det blevet almindeligt at bruge $t-$ teststørrelsen men så vurderet i den asymptotiske fordeling (altså den sande fordeling) under nul-hypotesen $\gamma = 0$:

$$t_\gamma \xrightarrow{F} \frac{\int_0^1 W(r) \mathrm{d} W(r)}{\{\int_0^1 W(r)^2 \mathrm{d} r\}^{1/2}}.$$

Dickey-Fuller-μ testen

- Antag modellen er $x_t = \mu + \phi x_{t-1} + w_t, \quad t \in \mathbb{N}$.
- Betragt $\mathrm{H}_0 : \phi = 1$. Formelt set, definerer dette en *random walk med drift*.
- Dickey & Fuller bruger den transformerede model

$$\nabla x_t = \mu + \gamma x_{t-1} + w_t,$$

med $\mathrm{H}_0 : \gamma = 0$. Dickey-Fuller (DF-$\mu$) teststørrelsen $t_\gamma^{(\mu)}$ er t-værdien for koefficienten γ i ovenstående regressionsligning.

Fordelingen af DF-μ teststørrelsen

- $t_\gamma^{(\mu)}$ er ikke t-fordelt eller normalfordelt under H_0. Den er ikke engang fordelt ligesom DF-0 teststørrelsen. I stedet har man

  $$n\hat{\gamma} \xrightarrow{F} \frac{\int_0^1 W_1(r)\mathrm{d}W_1(r)}{\int_0^1 W_1(r)^2\mathrm{d}r},$$

  under antagelse af, at $\gamma = 0$ og $\mu = 0$, og tilsvarende

  $$t_\gamma^{(\mu)} \xrightarrow{F} \frac{\int_0^1 W_1(r)\mathrm{d}W_1(r)}{\{\int_0^1 W_1(r)^2\mathrm{d}r\}^{1/2}},$$

  hvor $W_1(t)$ er en middelværdi-korrigeret Brownsk bevægelse. Den asymptotiske fordeling er påvirket af, at man har inkluderet en konstant i regressionen.

Dickey-Fuller-τ testen

- Antag nu at modellen er $x_t = \mu + \tau t + \phi x_{t-1} + w_t, \quad t \in \mathbb{N}$.
- Betragt $\mathrm{H}_0 : \phi = 1$. Formelt set, definerer dette en *random walk med drift plus lineær trend*.
- Dickey & Fuller bruger den transformerede model

$$\nabla x_t = \mu + \tau t + \gamma x_{t-1} + w_t,$$

med $\mathrm{H}_0 : \gamma = 0$. Dickey-Fuller (DF-$\tau$) teststørrelsen $t_\gamma^{(\tau)}$ er t-værdien for koefficienten γ i ovenstående regressionsligning.

Fordelingen af DF-τ teststørrelsen

- $t_\gamma^{(\tau)}$ er igen ikke t-fordelt eller normalfordelt under H_0. Den er heller ikke fordelt ligesom DF-0 eller DF-μ teststørrelsen. I stedet har man

$$n\hat{\gamma} \xrightarrow{F} \frac{\int_0^1 W_2(r)\mathrm{d}W_2(r)}{\int_0^1 W_2(r)^2\mathrm{d}r},$$

under antagelse af, at $\gamma = 0$, $\mu = 0$ og $\tau = 0$, og tilsvarende

$$t_\gamma^{(\tau)} \xrightarrow{F} \frac{\int_0^1 W_2(r)\mathrm{d}W_2(r)}{\{\int_0^1 W_2(r)^2\mathrm{d}r\}^{1/2}},$$

hvor $W_2(t)$ er en trend-korrigeret Brownsk bevægelse. Den asymptotiske fordeling er påvirket af, at man har inkluderet en konstant i regressionen.

Hvilken af Dickey-Fuller testene skal man bruge

- DF-0 testen bruges sjældent.
- DF-μ testen tester nulhypotesen om en random walk uden drift mod alternativhypotesen at tidsrækken er en stationær autoregression. Denne test er tilstrækkelig for ikke-trendede data (altså ikke deterministisk-trendede data).
- DF-τ testen tester nulhypotesen om en random walk muligvis med drift mod alternativet, at det er en trend-stationær proces. Denne test er tilstrækkelig for trendede data (altså deterministisk-trendede data).
- Alle kritiske værdier og p-værdier fås ud fra specielle tabeller af simulerede fraktiler.

DF-testene er følsomme overfor autokorrelation

- Ofte vil residualerne i regressionen

$$\nabla x_t = \mu + \tau t + \gamma x_{t-1} + w_t$$

  være autokorrelerede. AR(1) modellen er simpelthen for simpel. Dette påvirker fordelingen af t_γ under nulhypotesen. Der er to mulige løsninger:

  1. Brug en model som er tilstrækkelig for højere ordens autoregressioner: augmented DF-test, eller ADF test, som det kaldes.
  2. Juster t-værdien ved hjælp af en ikke-parametrisk korrektion, som opfanger den residuelle autokorrelation: Phillips & Perron testen.

- Vi fokuserer på den første af ovenstående.

Transformation af en autoregression

- Enhver AR(p) model

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t$$

kan omskrives på formen

$$\nabla x_t = \gamma x_{t-1} + \sum_{j=1}^{p-1} \tilde{\gamma}_j \nabla x_{t-j} + w_t.$$

- Der er en entydig korrespondance (en bijektion) mellem parametrene $\phi_1, \dots, \phi_p$ og $\gamma, \tilde{\gamma}_1, \dots, \tilde{\gamma}_{p-1}$.

Fordelene ved denne *tilde* repræsentation

- Hvis $\gamma = 0$ i *tilde* repræsentationen, så har AR polynomiet en enhedsrod (unit root). Bemærk, at $-\gamma = 1 - \phi_1 - \cdots - \phi_p$. Af denne grund er $\gamma = 0$ ækvivalent med I(1) tilfældet, forudsat der ikke er andre enhedsrødder. Test for $\gamma = 0$ giver en såkaldt unit root test.
- Dickey & Fuller viste at t_γ i *tilde* regressionen asymptotisk har DF-0 fordelingen, forudsat den sande model er AR(p). Det samme gælder for versioner med deterministiske led (DF-μ og DF-τ).

Hvordan bestemmer man *augmentation* p i praksis

Redundante (overflødige) lags af ∇x_t gør ikke fordelingen af DF teststørrelsen forkert, men det gør autokorrelation. Alternativer:

1. p fastsættes som en funktion af stikprøvestørrelsen, for eksempel $n^{1/3}$.
2. p fastsættes så der ikke længere er autokorrelation i residualerne.
3. p fastsættes vha. informationskriterier, AIC og/eller BIC.

Philips-Perron testen

Testen af Phillips & Perron konstrueres vha. to modifikationer af den (ikke-augmenterede) DF-μ eller DF-τ teststørrelser t_γ

1. t_γ erstattes af $t_\gamma \frac{\sigma_w}{\sigma_l}$, hvor σ_w estimerer den residuelle standardafvigelse og σ_l er kvadratroden af spektraltætheden i nul af residualerne.
2. Et additivt korrektionsled, som afhænger af observationerne og af σ_w og σ_l, fratrækkes t_γ.

Beregning af σ_l er lidt kompliceret, og testen har den samme fordeling som ADF testen.