Aikasarjoista päätöksiin: ennustaminen taloustieteissä

Operaatiotutkimus ja tilastotiede vakuutus- ja rahoitusalalla

Lauri Viitasaari, Aalto-yliopisto

Claude Opus 5.5

Esityksen kulku

1 **Aikasarjat ja päätöksenteko**
Ala tuottaa aikasarjoja, ja päätökset vaativat ennusteen ja sen epävarmuuden

2 **Neljä mallinnettavaa ilmiötä**
Riippuvuudet, volatiliteetti, muisti ja trendit simuloidussa ja todellisessa datassa

3 **Klassiset mallit ja tekoäly**
ARMA-oletukset, ilmiöiden matemaattinen kuvaus, klassiset ratkaisut ja tekoälymallit

4 **Tutkimus Aalto-yliopistossa**
SDE-mallit, äärellinen otos, raja-arvolauseet, ääriarvot, mallivirhe ja sovellukset

OSIO 1

Aikasarjat ja päätöksenteko

Ala tuottaa aikasarjoja, ja päätökset vaativat ennusteen ja sen epävarmuuden

Vakuutus- ja rahoitusalan keskeinen data on aikasarjoja



Osakkeet ja indeksit

Päivittäiset tuotot ja volatiliteetti



Korot ja tuottokäyrät

Diskonttaus ja vastuuvelan arvostus



Vahinkokorvaukset

Vahinkojen määrä ja suuruus ajassa



Kuolevuus

Pitkäikäisyysriski henkivakuutuksessa



Sää ja luonnonilmiöt

Myrsky-, tulva- ja satovahingot



Valuutat ja hyödykkeet

Kurssi- ja hintariskit

Hyvä päätös vaatii sekä ennusteen että tiedon sen epävarmuudesta



Epävarmuus kulkee koko ketjun läpi: datan, mallin ja parametrien virheet päätyvät päätökseen

Operaatiotutkimus vastaa kysymykseen, mitä pitäisi tehdä. **Tilastotiede** kertoo, kuinka varmoja voimme olla.

Ennustaminen ja mallintaminen ovat eri tehtäviä, ja päätöksenteko tarvitsee molempia

Ennustaminen

Mitä tapahtuu seuraavaksi?

- Onnistumista mitataan ennustevirheellä uudessa datassa
- Musta laatikko voi riittää, jos se ennustaa hyvin
- Esimerkki: huomisen volatiliteetti tai ensi vuoden korvausmeno

Mallintaminen

Miksi ja miten data syntyy?

- Onnistumista mitataan tulkittavuudella ja teoreettisilla takuilla
- Mahdollistaa skenaariot, stressitestit ja syy-yhteyksien arvioinnin
- Esimerkki: korkomalli

OSIO 2

Neljä mallinnettavaa ilmiötä

Riippuvuudet, volatiliteetti, muisti ja trendit simuloidussa ja todellisessa datassa

Neljä ilmiötä, jotka rahoitus- ja vakuutusdatan mallin on otettava huomioon

1

Spatiaaliset riippuvuudet

Sarjat liikkuvat yhdessä alueittain ja markkinoittain

Esimerkki

Myrskyvahingot naapurialueilla

2

Muuttuva volatiliteetti

Hajonta vaihtelee, ja suuret heilahtelut kasautuvat

Esimerkki

Kriisijaksot osake-markkinoilla

3

Muistiefektit

Menneet shokit vaikuttavat pitkään

Esimerkki

Volatiliteetin ja korkojen pysyvyys

4

Monimutkaiset trendit

Rakennemuutokset, kausivaihtelu, epälineaarisuus

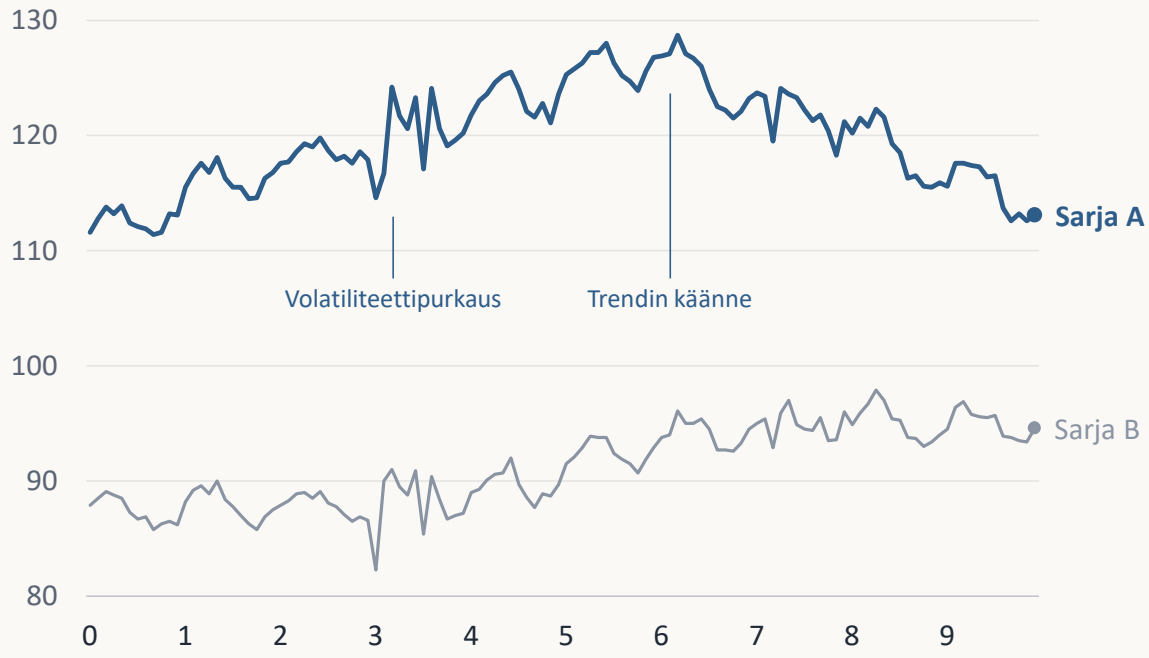
Esimerkki

Kuolevuuden aleneminen

Simuloitu esimerkki: kaksi sarjaa, joissa kaikki neljä ilmiötä näkyvät selvästi

Kaksi simuloitua sarjaa

Kuukausihavainnot, aika vuosina



Miten ilmiöt on rakennettu

1 Spatiaaliset riippuvuudet
Sarjojen shokit korreloivat (ρ = 0,85)

2 Muuttuva volatiliteetti
Hajonta kymmenkertaistuu kahdesti ja vaimenee hitaasti

3 Muistiefektit
Fraktionaalinen integrointi, d = 0,4

4 Monimutkaiset trendit
Kausivaihtelu, trendin käänne ja tasonmuutos

OSIO 3

Klassiset mallit ja tekoäly

ARMA-oletukset, ilmiöiden matemaattinen kuvaus, klassiset ratkaisut ja tekoälymallit

ARMA-malli on tulkittava ja teoreettisesti hallittu, mutta sen oletukset ovat tiukat

ARMA(p, q)

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

Nykyarvo riippuu omista menneistä arvoistaan ja menneistä satunnaisshokeista ε.

Box ja Jenkins (1970) tekivät mallista käytännön työkalun: tunnistus, estimointi ja diagnostiikka.

Vahvuudet

- Parametrit ovat tulkittavia
- Estimaattorien teoria tunnetaan hyvin
- Toimii myös lyhyillä sarjoilla

Rajoitukset

- Olettaa stationaarisuuden
- Lineaarinen ja vakiovarianssinen
- Lyhyt muisti: korrelaatio vaimenee eksponentiaalisesti

Kukin ilmiö rikkoo yhden ARMA-oletuksen ja vaatii oman matemaattisen rakenteen

Ilmiö	ARMA olettaa	Matemaattinen kuvaus ilmiölle
Spatiaaliset riippuvuudet	Yksi sarja kerrallaan	Vektoriprosessi $X_t \in \mathbb{R}^d$ ja ristikovarianssi $\Gamma(h) = \mathrm{Cov}(X_t, X_{t+h})$
Muuttuva volatiliteetti	Vakio ehdollinen varianssi	$X_t = \sigma_t \varepsilon_t$, jossa $\sigma_t^2 = \mathrm{Var}(X_t \mid \mathcal{F}_{t-1})$ on itse satunnainen prosessi
Muistiefektit	Autokorrelaatio vaimenee nopeasti	$\gamma(h) \sim C\, h^{2d-1}$, kun $h \to \infty$, joten $\Sigma \lvert\gamma(h)\rvert = \infty$
Monimutkaiset trendit	Vakio odotusarvo	$X_t = \mu(t) + s(t) + Y_t$: trendi, kausivaihtelu ja stationaarinen osa

Klassinen ratkaisu: kullekin ilmiölle oma malliperhe, jonka teoria tunnetaan hyvin

Ilmiö	Klassinen malli	Perusyhtälö
Spatiaaliset riippuvuudet	VAR- ja VARMA-mallit, spatiaaliset mallit	$X_t = A_1X_{t-1} + \ldots + A_pX_{t-p} + \varepsilon_t$
Muuttuva volatiliteetti	ARCH/GARCH, stokastinen volatiliteetti	$\sigma_t^2 = \omega + \alpha\, X_{t-1}^2 + \beta\, \sigma_{t-1}^2$
Muistiefektit	ARFIMA, fraktionaalinen Brownin liike	$(1 - B)^d\, \phi(B)\, X_t = \theta(B)\, \varepsilon_t$
Monimutkaiset trendit	Trendi- ja kausihajotelma, rakennekatkot	$X_t = \mu(t) + s(t) + Y_t$, katkot $\mu(t)$:ssä
Teoreettiset takuut	Vahvat: estimaattorien jakaumat tunnetaan	Kukin ilmiö vaatii oman laajennuksen

Tekoälymallit oppivat riippuvuudet suoraan datasta, mutta niiltä puuttuvat teoreettiset takuut

Rekurrentit verkot

LSTM, GRU

Muisti kulkee piilotilassa askeleesta toiseen

Transformerit

Perustusmallit, esim. Chronos ja TimeGPT

Esikoulutettu suurella määrällä eri sarjoja

Graafineuroverkot

GNN-ennustemallit

Sarjojen väliset riippuvuudet verkkona

Vahvuudet

- Epälineaarisuus ja joustavuus
- Tuhansia sarjoja yhdellä mallilla

Heikkoudet

- Paljon dataa vaativia ja vaikeasti tulkittavia
- Epävarmuuden arviointi usein puutteellista

OSIO 4

Tutkimus Aalto-yliopistossa

SDE-mallit, äärellinen otos, raja-arvolauseet, ääriarvot, mallivirhe ja sovellukset

Tutkimme jatkuvan ajan SDE-malleja, joihin kaikki neljä ilmiötä mahtuvat

$$dX_t = \mu(t, X_t)\,dt + \sigma_t\,dW_t$$

Perusmalli: driftti μ kuvaa trendin, diffuusio σ satunnaisen vaihtelun

1 Spatiaaliset riippuvuudet

$X_t \in \mathbb{R}^d$
$\mathrm{Cov}(dW_t) = \Sigma\,dt$

Sarjat jakavat korreloidun kohinan

2 Muuttuva volatiliteetti

$dv_t = \kappa(\theta - v_t)\,dt + \xi\sqrt{v_t}\,dZ_t$

Volatiliteetti $\sigma_t^2 = v_t$ on oma prosessinsa

3 Muistiefektit

$W_t \to B_t^H,\ H > 1/2$

Fraktionaalinen Brownin liike tuo pitkän muistin

4 Monimutkaiset trendit

$\mu(t, x)$ riippuu ajasta

Trendi, kausivaihtelu ja katkot driftissä

Miksi jatkuva aika: havaintovälit voivat olla epäsäännöllisiä, sama malli toimii eri aikaskaaloilla, ja AR- ja GARCH-mallit saadaan diskretointeina.

Tutkimme äärellisen otoksen rajoja, jotka antavat estimointivirheelle todennäköisyystakuun

- Asymptoottinen tulos kertoo, mitä tapahtuu kun n → ∞, mutta ei kuinka suuri virhe on tällä datalla
- Esimerkiksi 30 vuoden kuukausiaineistossa on vain 360 havaintoa, ja usein se on kaikki, mitä yhdestä riskistä on mitattu
- Äärellisen otoksen raja antaa virheelle todennäköisyystakuun:

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \delta(n, \varepsilon)$$

- Aikasarjan riippuvuus pienentää efektiivistä otoskokoa
- Tekoälymallin parametreja voi olla moninkertaisesti havaintoihin nähden

Tutkimme raja-arvolauseita: muisti ratkaisee, onko estimointivirhe normaalijakautunut

Keskeinen raja-arvolause

Lyhyt muisti

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, \sigma^2)$$

- Normaali rajajakauma, nopeus $n^{1/2}$
- Luottamusvälit tavalliseen tapaan
- Klassinen tilastollinen päättely pätee

Ei-keskeinen raja-arvolause

Pitkä muisti, H > 3/4

$$n^{2-2H}(\hat{\theta}_n - \theta) \to F_\infty$$

- Hitaampi nopeus, ei-normaali raja
- $n^{1/2}$-oletus tuottaa liian kapeat välit
- Päättelyyn tarvitaan oma teoriansa

Esimerkki: varianssin ja volatiliteetin kaltaiset neliölliset estimaattorit fraktionaalisen Brownin liikkeen mallissa. Raja-arvolause kertoo, minkä levyinen ja muotoinen luottamusväli on.

Tutkimme ääriarvoja: riskienhallinnan päätökset ratkeavat jakauman hännässä

Maksimit

Fisher–Tippett–Gnedenko

$$\max X_i \text{ normeerattuna} \rightarrow \mathrm{GEV}(\xi)$$

- Häntäindeksi $\xi > 0$: paksu häntä
- Ennustaa esim. vuoden suurimman tappion
- Toistumisaika: 99,5 % = kerran 200 vuodessa

Kynnyksen ylitykset

Pickands–Balkema–de Haan

$$P(X - u > y \mid X > u) \approx (1 + \xi y/\beta)^{-1/\xi}$$

- Ylitykset seuraavat GPD-jakaumaa
- Antaa VaR- ja ES-arviot suoraan
- Käyttää kaiken hännän datan

Aikasarjoissa: volatiliteetin kasautuminen tuo ääriarvot ryppäinä. Ekstremaali-indeksi $\theta \in (0, 1]$ kertoo ryppään koon, ja riippumattomuusoletus aliarvioi riskiä.

Kvantifioimme, miten sovitus- ja spesifiointivirhe näkyy virheellisinä päätöksinä

Sovitus- ja spesifiointivirhe	Virhe riskiluvussa	Päätöksen kustannus
Vakavaraisuus	**Hinnoittelu**	**Varaukset**
Solvenssi II:n pääomavaatimus perustuu 99,5 %:n VaR-arvoon yhden vuoden aikavälillä.	Aliarvioitu volatiliteetti johtaa alihinnoiteltuihin takuisiin ja optioihin.	Kuolevuustrendin virhe kertautuu pitkissä eläke- ja henkivakuutusvastuissa.
Häntäarvion virhe muuttaa suoraan sidottua pääomaa.	Yliarvioitu volatiliteetti heikentää kilpailukykyä.	Pienikin trendivirhe kasvaa vuosikymmenten aikana.

Samat menetelmät käytössä: sähkömarkkinat, metsänhoito ja korkomallit



Sähkömarkkinat

Datakeskusten rakentaminen muuttaa sähkön kysyntää ja hintadynamiikkaa

Ilmiöt
Hintapiikit, alueelliset riippuvuudet, kysynnän murros

Merkitys alalle
Energiariskit ja infrastruktuurisijoitukset



Metsänhoito

Hakkuu- ja hoitopäätösten optimointi tekoälymalleilla vuosikymmenten horisontilla

Ilmiöt
Pitkät trendit, puun hinnan ja kasvun epävarmuus

Merkitys alalle
Metsäomaisuus sijoituskohteena, ilmastoriski



Korkomallit

Korkojen dynamiikka jatkuvassa ajassa: $dr_t = \kappa(\theta - r_t)\,dt + \sigma\,dW_t$

Ilmiöt
Muisti, muuttuva volatiliteetti, korkoregiimit

Merkitys alalle
Vastuuvelka ja tasehallinta

Yhteenveto

1. Rahoitus- ja vakuutusdatassa on **neljä ilmiötä**, ja päätökset tarvitsevat ennusteen **epävarmuuden**.

2. Klassiset mallit antavat **takuut**, tekoälymallit **joustavuuden**. Kumpikaan ei yksin riitä.

3. Tutkimuksemme kertoo, **kuinka paljon ennusteeseen voi luottaa**: äärellisen otoksen takuut, raja-arvolauseet ja ääriarvot.

4. Mallivirhe voidaan muuttaa **päätösten kustannukseksi** pääomassa, hinnoittelussa ja varauksissa.

Kiitos! Kysymyksiä?

Yliopisto tuo teorian ja virheanalyysin, teollisuus datan ja todelliset päätösongelmat.

Lauri.viitasaari@aalto.fi

Yhteistyön muotoja

- Opinnäytetyöt ja väitöskirjat
- Yhteiset tutkimusprojektit
- Mallien validointi ja virheanalyysi