

Aikasarja-analyysiä taloudellisilla aineistoilla

Leena Kalliovirta, Luonnonvarakeskus

Leena.kalliovirta@luke.fi

Kurssi *Tilastotiede tutuksi*

HY matematiikan ja tilastotieteen laitos

19.11.2015

Sisältö

- Taustani tilastotieteilijänä
- Regressiomalli
- Aikasarja-analyysiä
 - Epälineaariset aikasarjamallit
 - Osa 1 A: Korkeero ja valuuttakurssi Euroalue - USA
 - Osa 1 B: Tulonjako teollisuusmaissa
 - Dominoivat trendit
 - Osa 2: Suomen raakapuumarkkinat

Taustani

- VTM, VTL ja VTT, Helsingin yliopisto
 - Pääaine tilastotiede
 - Sivuaineet matematiikka ja taloustiede
- Työpaikat:
 - Tutkijatohtori, HY
 - Ekonometrikko, Luke

Luonnonvarakeskus Luke

- Valtion tutkimuslaitos
 - Toimii Maa- ja metsätalousministeriön alaisuudessa
 - Yksityiskohtaisemmat tiedot löytyvät prof. Juha Heikkisen kalvoista (24.9.2015)
- Ekonometrikon tehtävät
 - Tutkimusta yhteistyössä taloustieteilijöiden kanssa
 - metsä- ja maatalouteen liittyvä empiirinen taloustiede
 - Menetelmien opetusta Luken henkilökunnalle
 - Menetelmätuki

Tutkimustyö

- Tutkimustöissä olen tarvinnut tietoja mm.
 - Tilastotieteen opinnoista
 - Tilastollinen päättely
 - Moniulotteinen aikasarja-analyysi
 - Epästationaarinen aikasarja-analyysi
 - Matematiikan opinnoista
 - lineaarialgebra
 - analyysi, mittateoria, stokastiikka
 - Taloustieteen opinnoista
 - Mikro- ja makrotaloustiede

Regressiomallit - Pikakurssi

Regressiomalli - havaintoaineisto

$$y_1, y_2, \dots, y_t, \dots, y_T$$
$$X_1, X_2, \dots, X_t, \dots, X_T$$

- Otoskoko T
- $y_t, t=1, \dots, T$ *-kiinnostava muuttuja*
- $X_t, t=1, \dots, T$ *-mahdollisesti selittää y_t :n arvoja*

Regressiomalli - havaintoaineisto

$$y_1, y_2, \dots, y_t, \dots, y_T$$
$$X_1, X_2, \dots, X_t, \dots, X_T$$

- Esimerkkeinä:
 1. Euroalueen ja USA:n korkojen välinen korkoero ja Euron ja US Dollarin valuuttakurssi
 2. Tulonjako: suurituloisimman 1%:n osuus kokonaistuloista, 6 teollisuusmaata
 3. Raakapuumarkkinat: Suomen mänty-, kuusi- ja koivukuitupuiden kantohintoihin vaikuttavat tekijät (mm. vientitullit ja verotuksen muutos)

Regressiomalli - malliyhtälö

*Selitettävän muuttujan vaihtelu
= selitetty osa + selittämätön osa*

$$y_t = X_t\beta + \varepsilon_t$$

- y_t selitettävä muuttuja
- X_t selittävät muuttujat ja tuntemattomat parametrit β
(= selitetty vaihtelu)
- ε_t virhetermi (= selittämätön vaihtelu)

Regressiomalli - malliyhtälö

$$y_t = X_t\beta + \varepsilon_t$$

- **Idea:** Kiinnostavat hypoteesit voidaan kirjoittaa parametrimuodossa ja niiden paikkansapitävyyttä testata tilastollisin testein
 - H_0 : ”Muuttujat X_t eivät selitä muuttujan y_t vaihtelua”
 $\Leftrightarrow H_0: \beta = 0$
 - Tilastollisen testin avulla tehdään johtopäätös (hypoteesi hylätään tai jätetään voimaan).

Regressiomalli – oletukset, selittämätön osa

$$y_t = X_t\beta + \varepsilon_t$$

Selittämättömän osan satunnaisvaihtelusta tehdään oletuksia, jotta mallin parametrit saadaan estimoitua ja testien jakaumat johdettua.

- Regressiomallin virhetermistä ε_t oletetaan:
 - Sama todennäköisyysjakauma kaikilla t
 - Toisistaan riippumattomia

Regressiomalli – diagnostiikka, selittämätön osa

$$y_t = X_t\beta + \varepsilon_t$$

*Edellä mainittujen oletusten voimassaolo on tarkistettava diagnostisin **testein ja graafisin tarkasteluin** käyttäen estimoidun mallin residuaaleja (eli virhetermien ε_t empiirisiä vastineita)*

Regressiomalli – oletukset, selitetty osa

$$y_t = X_t\beta + \varepsilon_t$$

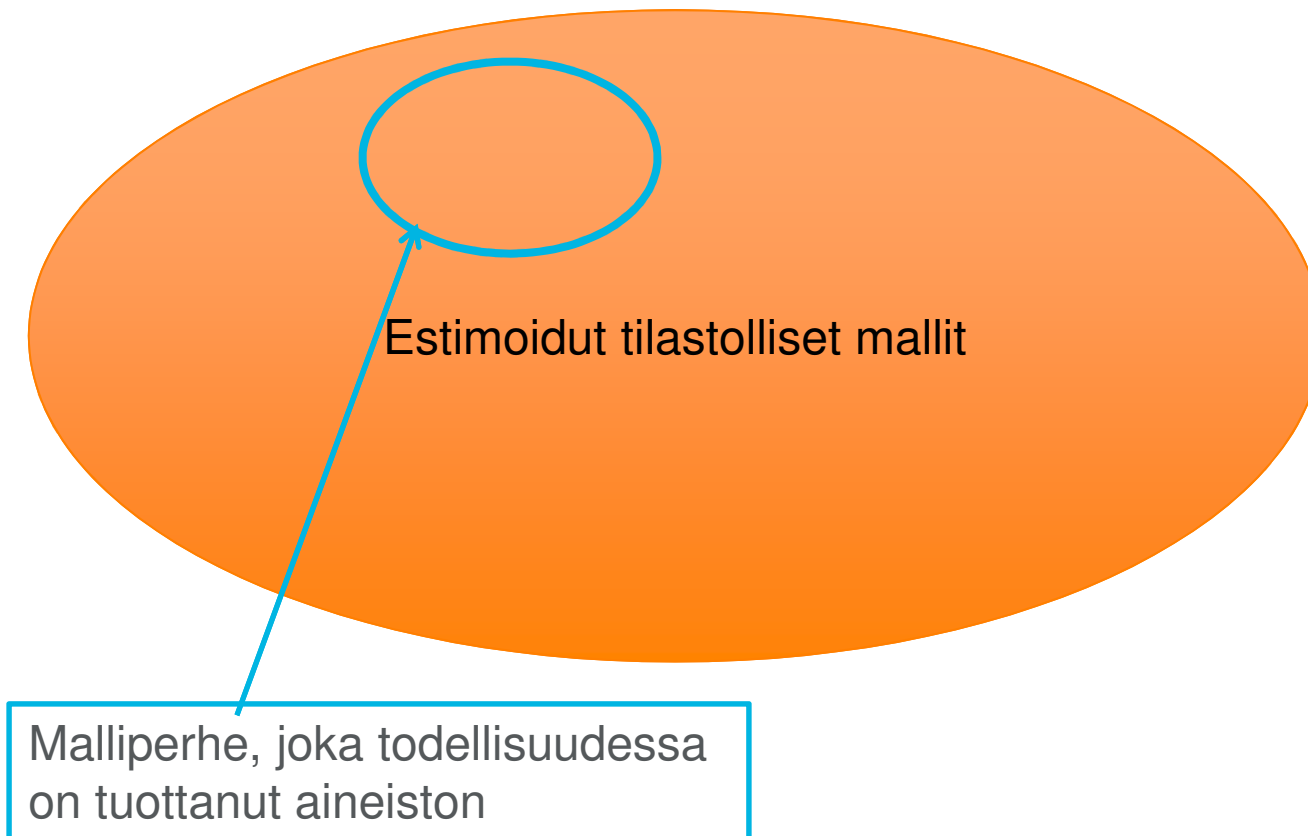
- Selitetyn osan kerroin β sama koko aineistossa
- Kaikki selittävät muuttujat ovat mallissa tarpeellisia
 - Kaikki β :n kertoimet ovat nolasta poikkeavia

Regressiomalli – diagnostiikka, selitetty osa

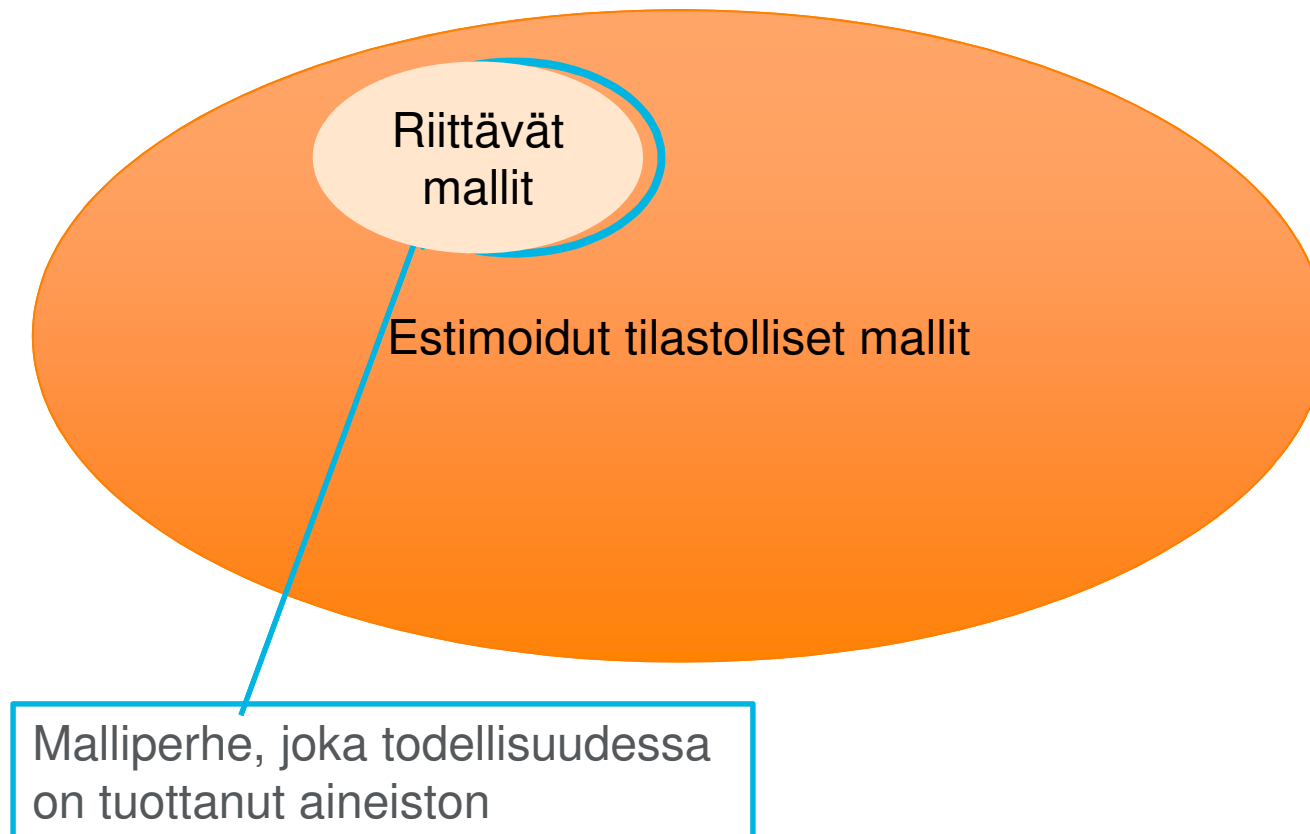
$$y_t = X_t\beta + \varepsilon_t$$

- Selitetyn osan vakioisuutta eli stabiilisuutta voidaan tutkia esim. estimoimalla β useasta osa-aineistosta
 - Graafiset menetelmät ja tilastolliset testit
- Kunkin selittävän muuttujan tarpeellisuutta voidaan tutkia tilastollisin testein
 - Tavoite: Säästeliäs malli, jossa on vain tilastollisesti merkitseviä kertoimia (parametreja).

Tilastollisen mallintamisen tavoite – Löytää aineiston tuottanut, *oikea* malli



Tilastollisen mallintamisen tavoite – Diagnostiikan läpäisevät mallit ovat riittäviä malleja



Aikasarjamallit - Esimerkkejä

Aikasarjamallit - regressiomallin yleistys

- **Havaintoaineiston ominaisuudet määrittelevät käytettävän tilastollisen mallin!**
- Talouden aikasarja-aineistoissa on tyypillisesti voimakas autokorrelaatio eli ajan suhteen peräkkäiset havainnot ovat voimakkaasti korreloituneita
- Tavanomaisen regressiomallin oletukset eivät päde!

VAR - malli - regressiomallin yleistys

VAR-malli

$$y_t = A_0 + A_1 y_{t-1} + \varepsilon_t$$

sallii ***selitettävien muuttujien***

- ***samanaikaiset ja***
- ***viivästetyt vaikutukset toisiinsa***

Aikasarjamallit - VAR mallien yleistyksiset

1) Epälineaariset mallit

- Rakenteelliset muutokset

2) Dominoivat trendit = Erittäin vahva autokorrelaatio

- Ns. yhteisintegraation mahdollisuus

OSA 1: Epälineaariset mallit

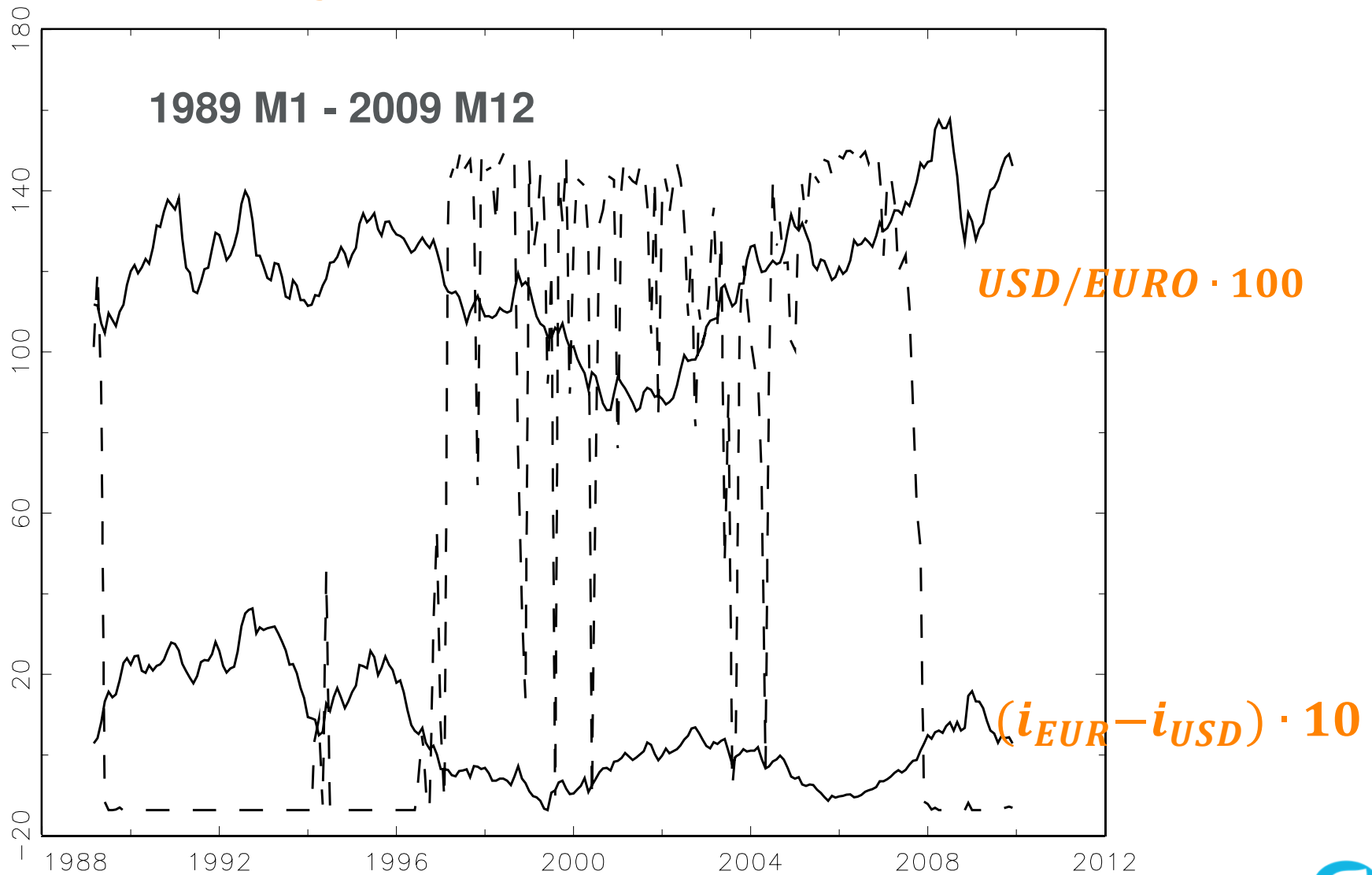
A: Korkeero ja valuuttakurssi

-

- 1) dynamiikan ymmärtäminen
- 2) ennustaminen

Lähde: *Gaussian Mixture Vector Autoregression*
Leena Kalliovirta, Mika Meitz ja Pentti Saikkonen

Korkoero ja valuuttakurssi - Aineisto



Korkoero ja valuuttakurssi - Mallit

- Lineaarinen VAR malli
 - Ei ole riittävä diagnostiikan perusteella
- Epälineaarinen GMVAR malli
 - On riittävä diagnostiikan mukaan

- GMVAR malli

$$y_t = \sum_{m=1}^2 s_{t,m} (A_{m,0} + A_{m,1} y_{t-1}) + \varepsilon_t$$

- Vectoriautoregressivisen (VAR) mallin yleistys
- ”Regiimien” avulla sallitaan rakenteellinen muutos

Korkoero ja valuuttakurssi - Tulokset

- Dynamiikka - Kaksi regiimiä
 - Matalan vaihtelun regiimi: Taso matalampi ja virhetermit korreloimattomia
 - Korkean vaihtelun regiimi: Taso korkeampi ja virhetermit positiivisesti korreloituneita
- Ennustekyvyn vertailu: 2010 M1 - 2013 M12
 - GMVAR malli ennustaa paremmin kuin aiemmat lineaariset tai epälineaariset mallit
 - MSPE
 - Ennusteiden jakauma

OSA 1: Epälineaariset mallit

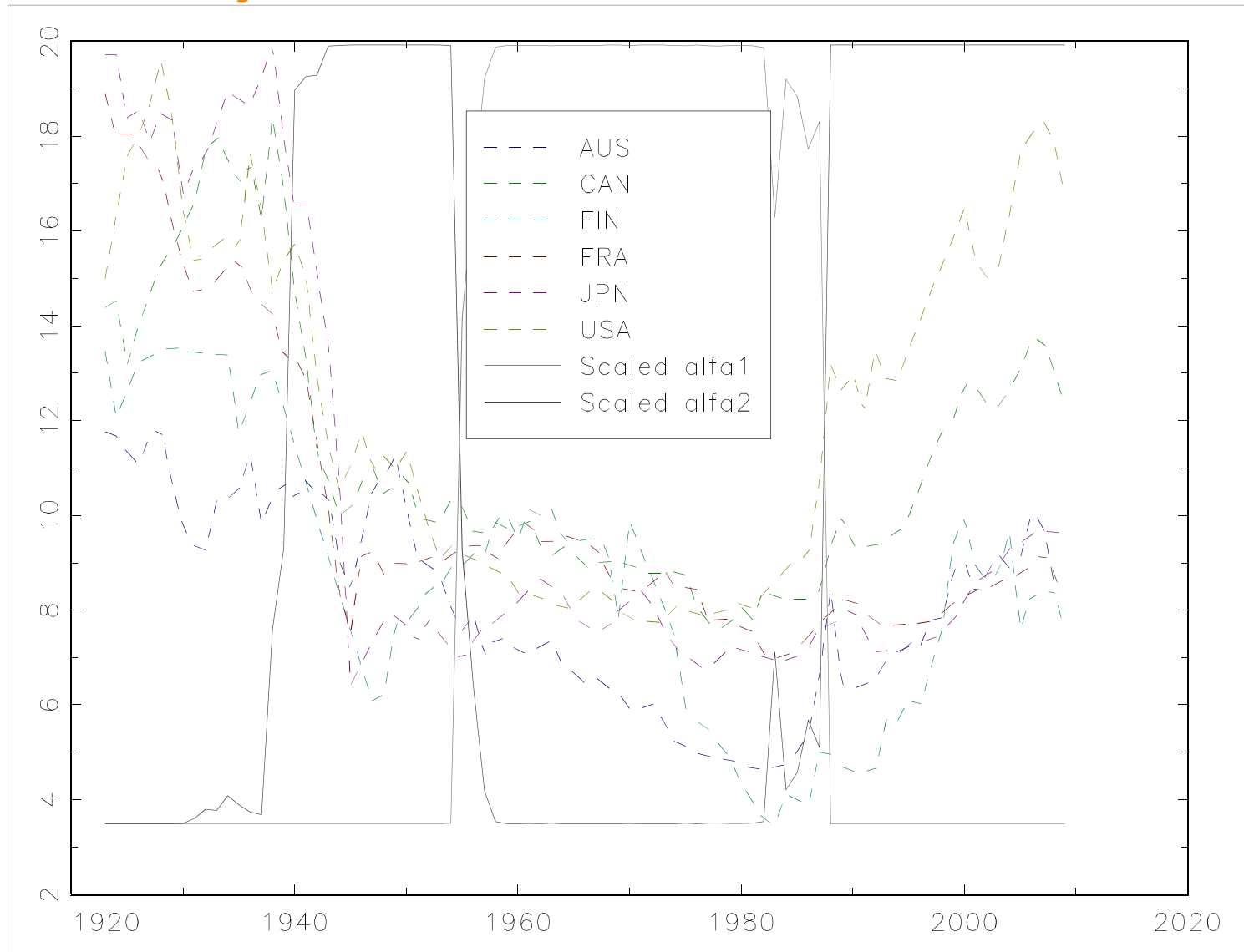
B: Tulonjakoaineisto

-

- 1) dynamiikan ymmärtäminen
- 2) vaikutukset maiden välillä

Lähde: *Nonlinearity and cross-sectional dependence of income inequality*
Tuomas Malinen ja Leena Kalliovirta

Tulonjako teollisuusmaissa - Aineisto



Tulonjako teollisuusmaissa - Malli

- GMVAR malli

$$y_t = \sum_{m=1}^2 s_{t,m} (A_{m,0} + A_{m,1} y_{t-1}) + \varepsilon_t$$

- ”Regiimien” avulla sallitaan rakenteellinen muutos
- Teoreettiset ominaisuudet tunnetaan:
 - On mahdollista tehdä tasojen ja vaihtelun tasojen vertailua regiimien välillä

Tulonjako teollisuusmaissa - Tulokset

- Dynamiikka - Kolme regiimiä
 - Yhteinen matalan vaihtelun regiimi: Taso matalampi ja virhetermit korreloimattomia
 - Yhteinen korkean vaihtelun regiimi: Taso korkeampi ja virhetermit positiivisesti korreloituneita
 - Ranskan ja Japanin osalta korkean vaihtelun regiimi jakautuu kahteen eri tasoiseen osaan
- Dynamiikka – USA:n tulonjako
 - Määrää yhteisen regiimin siirtymäajankohdat
 - Selittää muiden maiden tulonjakoa eli ”edeltää” muita
 - Muut maat seuraavat USA:n tulonjakoa

OSA 2: Dominoivat trendit

Kotimaan kuitupuumarkkinoiden
kantohinnat

-

Ulkoisten tekijöiden vaikutukset

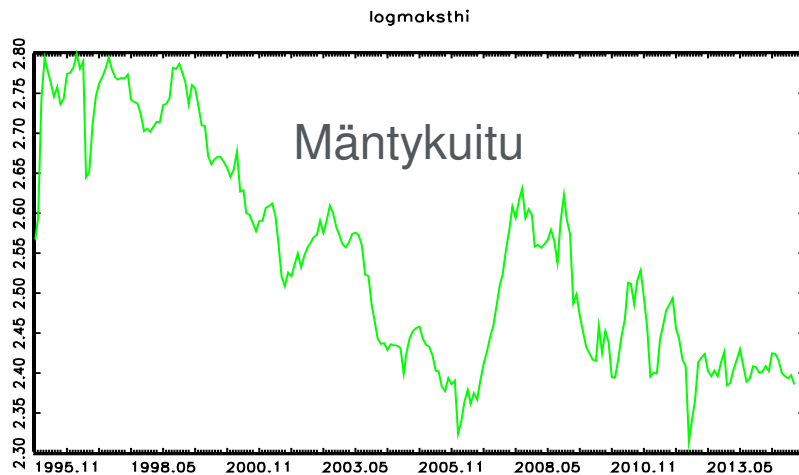
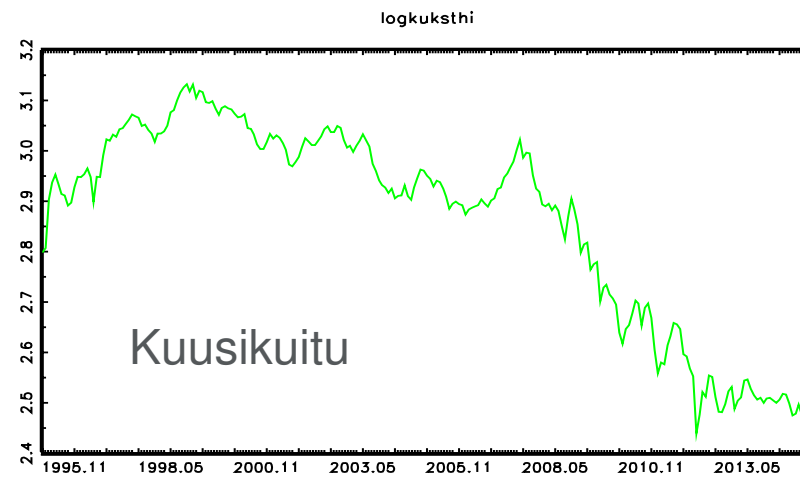
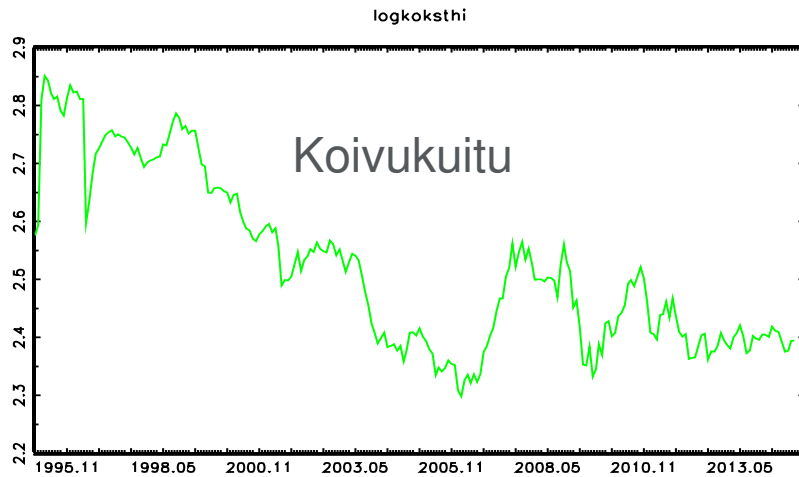
Työpäperi kesken - tekijät Leena Kalliovirta ja Riitta Hänninen

Aineisto

- Mallinnetaan tukkuhintaindeksillä (THI) deflatoituja hintoja.
 - Reaalihinnat (eli inflaation vaikutus huomioitu)
- Tarkasteluajanjakso 1995 M1- 2014 M10.

Kuitujen hinnat (selitettävät muuttujat)

Plot of Time Series 1995.01–2014.10, T=238



Aineiston ominaisuuksia

- *Kuukausittaiset* kuitupuiden hinnat riippuvat erittäin voimakkaasti omista menneistä hinnoistaan
- Sarjojen keskinäinen voimakas riippuvuus – yhteiset trendit eli ns. yhteisintegroituvuusrelaatiot
- VEC-malli on kehitetty juuri tällaisia aikasarjoja varten

VEC-mallin rakenne

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + BX_t + \varepsilon_t$$

- jakaa aineistossa olevat riippuvuudet
 1. pitkän aikavälin tasapainorelaatioon (dominoivat trendit) ja
 2. lyhyen aikavälin virheenkorjausmekanismiin, joka palauttaa pitkän aikavälin tasapainon.
- Regressiomallin yleistys
 - Selitettävät muuttujat ovat hintojen muutoksia (differenssejä)
 - Selittäviä muuttujia ovat poikkeamat tasapainorelaatiosta, viivästetyt differenssit ja eksogeeniset muuttujat

VEC-mallin rakenne

- Ulkopuoliset tekijät mukana ns. indikaattorimuuttujina
 - Verotuksen siirtymäaika (2005M12 asti)
 - Kartelliaika (1997M1-2004M4)
 - Useita Venäjän puutullien muutoksia
 - Työtaistelu (2005M5-M6)
- Kausivaihtelun indikaattorit
- Varmistetaan, että näiden tekijöiden mahdollinen vaikutus sallitaan mallissa.

Estimointitulokset: *VEC-malli*

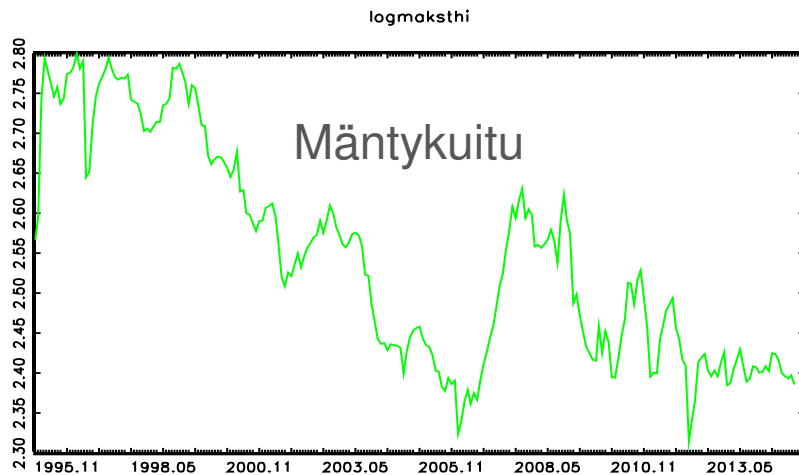
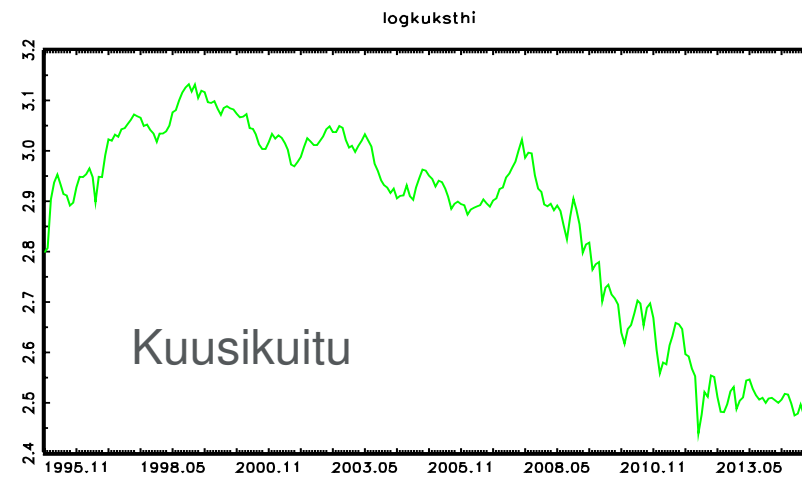
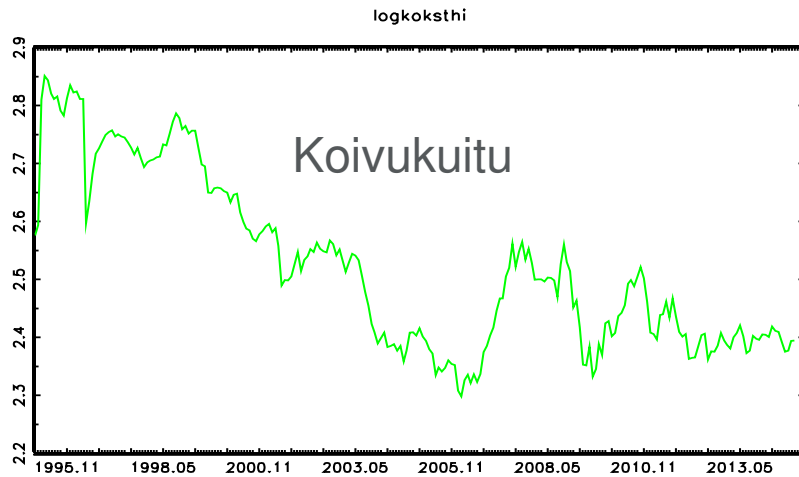
- **Rakenteellinen muutos:**
 - Vaihtelu kasvaa vuoden 2007 aikana.
 - Estimointiajanjakso lyhennetty 1995M2 – 2006M12
- **VEC-malli on riittävä malli**

Estimointitulokset: *VEC-malli*

- Tulokset:
 1. Hinnoilla on yhteinen trendi eli ne ovat yhteisintegroituneita
 2. Samanaikainen voimakas korrelaatio
 - Sama, likimain yksiulotteinen shokki generoi hintasarjat
 3. Osa ulkoisista tekijöistä on tilastollisesti merkitsevästi vaikuttanut hintoihin.

Kuitujen hinnat (selitettävät muuttujat)

Plot of Time Series 1995.01–2014.10, T=238



Tutkimusaihe jatkossa

Yhdistetään dominoivat trendit
ja GMVAR malli

-

Sovellus: raakapuumarkkinat
rakenteellinen muutos mallissa mukana

