

**Mat-2.177 Operaatiotutkimuksen
projektityöseminaari 2007**

**Loppuraportti
Korrelaatioiden ja varianssin estimointi
kiinteistöportfolion tuotolle**

27. 4. 2007

Kohdeorganisaatio:

Tapiola

Yhteyshenkilöt:

Ville Koskinen

Samu Anttila

Ryhmä:

Lauri Haapamäki

Tapani Hyvämäki (Projektipäällikkö)

Otso Manninen

Heikki Peura

Mirko Ruokokoski

1 Johdanto	3
1.1 Tavoitteet ja rajaukset	3
1.2 Raportin rakenne	4
2 Kirjallisuustutkimus.....	4
2.1 Mallinnustavat aiemmissä tutkimuksissa	6
3 Mallintamiseen käytetty data	8
3.1 Selitettävä muuttuja.....	9
3.2 Selittävät muuttujat	9
4 Kiinteistömarkkinoiden mallintaminen	10
4.1 Neuroverkko.....	11
4.1.1 Tulokset.....	12
4.2 ARIMAX	18
4.2.1 Tulokset.....	19
4.3 Garch-malli	23
4.3.1 Tulokset.....	23
5 Yhteenveto.....	32
6 Pohdintoja	33
Viitteet	34
Liitteet.....	35
Liite1: Neuroverkon matlab-koodi.	35
Liite 2: GARCH-malli	39

1 Johdanto

Tässä raportissa esitelty projekti on tehty kevään 2007 aikana Teknillisen korkeakoulun opintojaksolle Mat-2.177 Operaatiotutkimuksen projektityöseminaari. Työn asettajana toimi Tapiola-ryhmä, ja työn tavoitteena oli identifioida kiinteistöjen arvojen kehitykseen vaikuttavia tekijöitä sekä edelleen mallintaa kiinteistöportfolion tuottoa pyrkien yhteismitallistamiseen osakemarkkinoiden kanssa.

1.1 Tavoitteet ja rajaukset

Tapiolan asettama tehtävänanto oli etsiä menetelmiä, joiden avulla voidaan tuottaa parempia estimaatteja eri sijoituskohteiden välisen tuoton kovarianssimatriisille silloin, kun osa tarkasteltavista sijoituskohteista noteerataan vain korkeintaan muutamia kertoja vuodessa. Tarkoituksena oli siis tarkastella institutionaalisen sijoittajan vaihtoehtoja osakkeista ja kiinteistöistä koostuvan sijoitusportfolion arvioinnissa.

Osake- ja joukkovelkakirjamarkkinoilta tilastotietoja on saatavilla erittäin runsaasti. Tilanne on kuitenkin oleellisesti erilainen kiinteistömarkkinoiden kohdalla. Kiinteistöille ei ole saatavissa päivittäistä noteerausta (liikekiinteistöjen arvoa kuvaavia indeksejä on saatavilla yleensä vain vuosittain), eikä niillä ole selkeää yksiselitteistä markkina-arvoa paitsi myyntihetkellä. Monet sijoittajat arvioivat kiinteistöjen arvot vuosittain kassavirta-analyysille, eikä tämä ole välttämättä niiden todellinen käypä arvo. Apuna saatetaan myös käyttää indeksejä, jotka eivät ole nekään aivan yksinkertaisia: Yleensä tällainen indeksi pohjautuu edellisen vuoden todellisiin kiinteistökauppoihin, ja ne antavat jonkinlaisen painotetun keskiarvon tiettytyypiselle kiinteistöneliölle, tms. Toisaalta taas jos indeksi on laskettu alkuvuodesta, loppusyksystä usein joudutaan turvautumaan omiin varsin karkeisiin arvioihin kiinteistöjen hintojen kehityksestä. Nyt tavoitteena oli siis etsiä malleja, joilla kiinteistöjen arvoa voitaisiin kuvata paljon lyhyemmillä aikaväleillä mahdollisimman luotettavasti.

Toisaalta työssä oli lisäksi myös huomioitava, että saadun mallin pitäisi pystyä ennustamaan kiinteistöjen hintojen kehittymistä. Tämä asettaa huomattavia rajoituksia mahdollisesti

käytettävien parametrien suhteen, sillä selittävien muuttujien pitää olla jollain tapaa ennustettavissa.

Projektisuunnitelmassa tehtävä jaoteltiin neljään osaan:

- Kirjallisuustutkimus: etsitään erilaisia aikaisemmin tutkittuja menetelmiä ongelman ratkaisemiseksi
- Valitaan kirjallisuustutkimuksen perusteella muutama malli, joihin keskitytään.
- Mallien validointi: suoritetaan tilastollisia testejä saaduille tuloksille
- Optimoinnit estimoiduilla kovarianssimatriiseilla.

Näistä viimeinen päädyttiin projektin kuluessa hylkäämään aikataulullisista syistä. Sen sijaan keskityttiin mahdollisimman hyvien mallien luomiseen.

1.2 Raportin rakenne

Raportin luvussa 2 on esitelty lyhyesti tekemämme kirjallisuuskatsaus, jossa tarkastellaan tutkimuksia, jotka käsittelevät suoraan tai sivuavat joltain olennaiselta osalta aiheitamme. Pyrimme esittelemään tapoja, joilla näissä tutkimuksissa on käsitelty kiinteistömarkkinoiden mallintamista sekä esittelemään tärkeimpiä havaintoja sekä tuloksia. Luvussa 3 on käsitelty tutkimuksessa käytettyä aikasarjadataa ja luvussa 4 esitellään valitut mallit sekä näiden estimoinneista saadut tulokset, joita vertaillaan ja arvioidaan luvussa 5.

2 Kirjallisuustutkimus

Osake- ja kiinteistösijoitusten yhteismitallistamista koskevia tutkimuksia ei juuri ole olemassa. Yhteismitallistaminen tarkoittaisi ideaalitulanteessa sitä, että kiinteistösalkun arvo (tai hyvä arvio siitä) olisi saatavilla yhtä usein kuin osakesalkunkin. Tämä on käytännössä mahdotonta: osakkeiden arvoja noteerataan käytännössä jatkuvasti pörssissä, kun taas kiinteistöille ei ole vastaavaa tarkkaa arvon mittaria. Kiinteistöjen arvojen kehitystä seurataan yleisesti erilaisten indeksien avulla, joista useimmat lasketaan vain vuosittain.

Suurin osa alan tutkimuksesta on keskittynyt Yhdysvaltoihin, jonka markkinat ovat suuret, ja toisaalta indekseistä on saatavilla luotettavaa historiadataa pitkältä ajanjaksolta. Aikasarjan pituuden kannalta nimenomaan olennaisinta on, että markkinat ovat toimineet samalla periaatteella koko tutkittavan ajan. Pääosin vain lyhyen aikaa vapaina markkinoina toimimisen takia on Euroopassa alan tutkimusta tehty pääosin Isossa-Britanniassa – monessa pienemmässä maassa (Suomi mukaan lukien) markkinoiden pienempi koko ja pitkien yhtenevien aikasarjojen puute on vaikeuttanut tutkimuksen tekemistä. Tarvetta tutkimukselle selvästi kuitenkin on, sillä globalisoituvassakin maailmassa eri maiden markkinat eroavat mm. lainsäädännöllisistä sekä maantieteellisistä syistä huomattavasti toisistaan [Bon03].

Kiinteistöjen tuottoja koskevia tutkimuksia on olemassa runsaasti, niin teoreettisia kuin empiirisiäkin. Hyvin yleinen tutkimuskohde on verrata kiinteistöjen tuottoja ja riskiä muihin sijoitusmuotoihin, kuten osakemarkkinoihin. Tällaisista tutkimuksista on saatavilla varsin kirjavia tuloksia riippuen käytetyistä menetelmistä sekä datasta. Toinen suosittu tutkimuksen kohde on ollut kiinteistöjen optimaalinen osuus diversifoidussa sijoitusportfoliossa. Varsin yleistä on myös keskittyä vuokrien tutkimiseen kiinteistöjen arvonmuutoksen tai tuottojen asemesta, sillä tämä on huomattavasti helpompaa. Diskontatuilla vuokran kassavirta-analyysillä päästään ehkä hieman vähemmällä oletuksilla ja tulokset on paljon helpompi esittää läpinäkyvästi. Kaiken kaikkiaan tutkimus on viime aikoina siirtynyt yhä enemmän perinteisestä yksittäisten rakennusten arvottamiseen perustuvasta tutkimuksesta kokonaisten sijoitusportfolioiden arvojen mallintamiseen.

Monet tutkimuksista pohtivat varsin ansiokkaasti kiinteistömarkkinoihin ja kiinteistöjen tuottoihin vaikuttavia faktoreita. Näissä tutkimuksissa ollaan kuitenkin useimmiten kiinnostuneita pitkän aikavälin keskimääräisistä tuotoista ja tyydytään selittämään erilaisin mallein esimerkiksi jotain vuosittaista indeksiä. Esimerkiksi kiinteistöjen arvon kuukausittaista vaihtelua selittäviä tutkimuksia ei löydy.

Tutkimuksissa on havaittu, että kiinteistöjen tuotot korreloivat makroekonomisten tekijöiden kanssa. Korrelaatio näkyy tutkittaessa niin kiinteistösijoitusyhtiöiden kurssi-indeksejä (REIT, Real Estate Investment Trust), osakkeiden arvo-indeksejä (esim. Karolyi ja Sanders, [Kar98]) kuin kiinteistöjen heuristisiin arviointeihinkin perustuvia indeksejä (esim. Karakozova,

[Kar02]). Monet näistä makroekonomisista tekijöistä kuvaavat käytännössä talouden aktiivisuutta ja reagoivat usein samankaltaisesti suhdannevaihteluihin. Niinpä näiden keskinäiset korrelaatiot ovat verraten suuria, minkä takia kovin monen makroekonomisen tekijän käyttäminen selittäjänä ei välttämättä paranna tuloksia.

Yleisesti selittäjinä käytettäviä makroekonomisia tekijöitä ovat mm. bruttokansantuote (BKT), palvelusektorin tuotto ja erilaiset työllisyysasteet. Näistä BKT on usein havaittu parhaimmaksi selittäväksi tekijäksi. BKT kuitenkin lasketaan vain neljä kertaa vuodessa, joten sen soveltuvuus kuukausittaisen aikasarjan selittämiseen on kyseenalainen. Muita selittäviä tekijöitä tutkimuksissa ovat olleet mm. inflaatio, korkojen kehitys ja joukkovelkakirjojen tuotto sekä osakemarkkinoiden volatilititeetti.

Kiinteistöjen arvoa on pyritty joissain tutkimuksissa lähestymään myös niiden tarjontaa tutkimalla – esimerkiksi uusien toimistorakennusten valmistumisen kautta.

Suomessa kiinteistöjen arvojen tutkiminen on ollut verrattain vähäistä. Helsingin seudun toimistotuottoja on tutkittu jonkin verran, yhtenä Olga Karakozovan väitöskirja [Kar02], jossa pyritään mallintamaan Helsingin keskustan toimistojen vuokrien sekä tuottojen kehitystä vuosina 1971–2001 sekä niiden ennustamista tulevaisuudessa. Karakozova käyttää kiinteistöjen arvoa kuvaavana selitettävänä muuttujana vuosittaista KTI-indeksiä (kotimainen kiinteistöjen tuottoa kuvaava indeksi) ja päätyy löytämään yli 50 % selitysasteen dynaamisella regressiomallilla. Tutkimus antaa hyvän kuvan tärkeimmistä kiinteistöjen arvon muutoksen Suomen oloissa vaikuttavista tekijöistä (BKT kriittinen), vaikkakin sen luotettavuutta heikentävät tutkimusaikana markkinoiden rakenteessa tapahtuneet muutokset (markkinoiden vapautuminen).

2.1 Mallinnustavat aiemmissa tutkimuksissa

Kiinteistömarkkinoiden kuvaamiseen on tutkimuksissa käytetty monenlaisia mallinnustapoja. Suosituimpia näistä ovat olleet erilaiset regressiomallit. Tavallisessa regressiossa kuvataan selitettävän muuttujan lineaarista tilastollista riippuvuutta selitettävistä muuttujista. Usein selittävät muuttujat (esim. BKT, inflaatio) kuvaavat kiinteistöindeksien muutoksia vasta

vaikkapa vuoden viipeellä, joten monissa regressiomalleissa käytetään selittäjinä viivästettyjä muuttujia, ts. kyseessä on dynaaminen regressiomalli (jakautuneen viipymän malli).

Pelkkä jakautunut viipymä ei kuitenkaan välttämättä riitä selittämään muutoksia. Toimiva vaihtoehto on ollut tutkia ns. ARIMAX-malleja (integrated auto-regressive moving average model with exogenous explanatory variables), jossa siis viivästettyihin ulkoisiin selittäviin muuttujiin yhdistetään aikasarjan autoregressio sekä virhetermien liukuva keskiarvo (esim. [Kar02]). Toinen ARIMAXin kaltainen vaihtoehto on ns. virheenkorjausmalli, jossa käytetään samaan tapaan autoregressiota ja lisäksi ylimääräisenä selittävänä tekijänä mallin edellistä virhetermiä (jota siis ”korjataan”). Myös ARMA-malleja ilman ulkoisia selittäjiä on käytetty (esim. Chin ja Fan [Chi05]). Regressiomallit toimivat useimmiten varsin hyvin, kun halutaan selittää historiadataa, mutta niiden ennustuskyky on rajallinen.

Chin ja Fan käsittelevät tutkimuksessaan Singaporen markkinoista myös ARCH/GARCH-mallia (generalized autoregressive conditional heteroscedasticity), jota käytetään talouden tutkimuksessa yleisesti kuvaamaan aikasarjoja, joiden volatilitteetti muuttuu ajan myötä (ts. heteroskedastisia aikasarjoja). Nykyinen virhetermi selitetään aikaisempien virhetermien varianssien funktiona.

Eräissä tutkimuksissa kiinteistömarkkinoita on mallinnettu esimerkiksi ns. CAPM-mallilla (Capital Asset Pricing Model) sekä erilaisin faktorimallein. Nämä tutkimukset ovat kuitenkin usein käsitelleet varsin rajattuja erikoistapauksia markkinoista ja yleisluontoisempien tutkimusten puuttuessa on vaikea uskoa ko. mallien soveltuvan kovin hyvin tämän tutkimuksen tarkoituksiin.

Eräs vaihtoehto kiinteistömarkkinoiden mallintamiseen on neuroverkko. Rosenblatt sovelsi ensimmäisenä neuroverkkoja hahmontunnistukseen 1950-luvun loppupuolella. Kuitenkaan neuroverkkoja ei ole perinteisesti sovellettu taloudellisissa malleissa. Shachmurove ja Witkowska [Sha00] vertailivat kansainvälisen arvopaperimarkkinoiden informaatiota käyttämällä lineaarista regressiota ja neuroverkkoja. Vertailu osoitti, että neuroverkot ennustavat paremmin arvopaperin tuoton kuin perinteiset pienimmän neliösumman menetelmät ja yleiset regressiomallit. Connellan ja James [Con98] tutkivat neuroverkon

käyttöä ennustamaan kiinteistön reaalisoituarvoa. Yu [Yu99] käytti neuroverkkoa ennustamaan Nikkei-indeksiä. Tutkimuksessa osoittautui, että neuroverkko pärjäsikin paremmin kuin konventionaaliset ARIMA-mallit ennustamisessa.

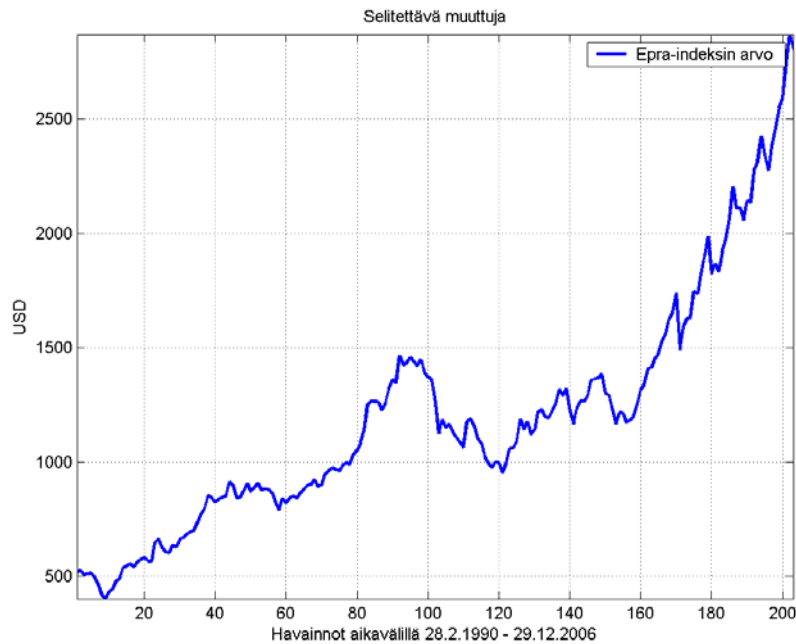
3 Mallintamiseen käytetty data

Suomessa Kiinteistötalouden instituutti ry laskee vuosittain ns. KTI-indeksin, joka mittaa suorien kiinteistösijoitusten kokonaistuottoa. Mukana on 22 suurta kiinteistösijoittajaa ja kaikkiaan noin 60 % kaikkien institutionaalisten sijoittajien sekä kiinteistösijoitusyhtiöiden omistamasta kiinteistökannasta. Tuotot perustuvat yksittäisistä kiinteistöistä koottuihin tuotto- ja kulutietoihin sekä kiinteistöjen arvioinnin ja vuokrauksen tietoihin. Vuosittaisena indeksinä KTI:n kiinteistöindeksi ei kuitenkaan soveltunut tämän tutkimuksen tarkoituksiin. Toisaalta Suomesta on nykyään olemassa myös jatkuvasti kiinteistösijoitusyhtiöiden arvoista päivittyvä EPRAn (European Public Real Estate Association) laskema kiinteistöindeksi. Tämän ongelmana on kuitenkin sen kattavuus: mukana on toistaiseksi vain muutamia yhtiöitä ja dataa on saatavilla vain muutamien vuosien ajalta.

Muualta maailmasta saatavat indeksit ovat kattavampia. Esimerkiksi Yhdysvalloissa suosittuja tutkimuksen kohteita ovat NAREIT-indeksi (The National Association of Real Estate Investment Trusts), joka lasketaan samaan tapaan kuin EPRAn luvut, kiinteistöjen arvioihin perustuva NCREIF (National Council of Real Estate Investment Fiduciaries) sekä keskimääräisiin kauppahintoihin perustuva NREI (National Real Estate Index). Lisäksi Yhdysvalloista on saatavilla pitkältä aikaväliltä monipuolista tilastotietoa erilaisista makrotaloudellisista kiinteistöjen arvojen muutoksista mahdollisesti selittävästä tekijöistä.

Koska Yhdysvalloista toisin kuin Suomesta on siis saatavana riittävästi havaintoaineistoa tilastollisesti merkitsevän mallin luomiseen, käytetty data on USA:n kiinteistömarkkinoilta. Havainnot ovat kuukauden välein aikavälillä 28.2.1990 - 29.12.2006.

3.1 Selitettävä muuttuja



Kuva 1 USA:n kiinteistömarkkinoita kuvaava indeksi kuukausittain (EPRA-NAREIT).

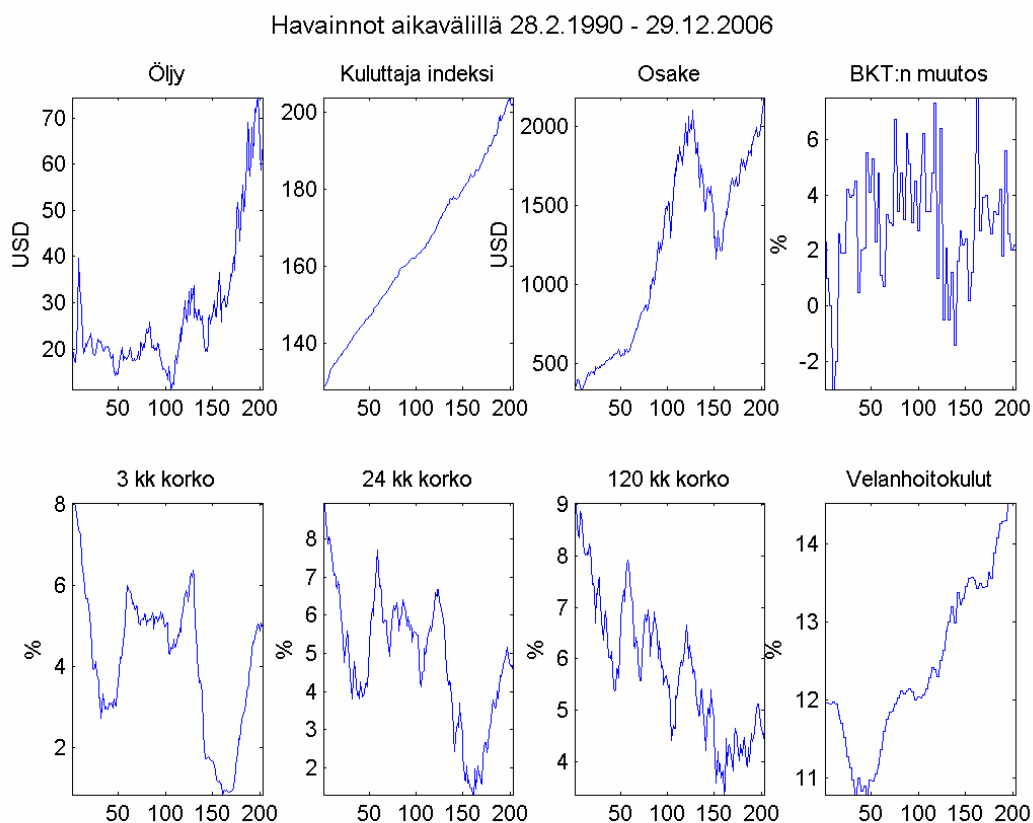
Selitettävänä muuttujana on USA:n kiinteistömarkkinan arvoa kuvaava indeksi, USA:n dollareina. Indeksi on peräisin EPRA:lta (European Public Real Estate Association, www.epra.com). Kuvassa 1 on tästä indeksistä havainnollistus vuodesta 1990 eteenpäin vuoden 2006 loppuun. Kuvasta voidaan nähdä selvä eksponentiaalinen käyttäytyminen indeksin arvoissa, varsinkin vuoden 2003 jälkeen (havainnoista 156 eteenpäin). Tämä viittaa siihen, että ennen mallien luontia dataa on esikäsiteltävä, esimerkiksi differensoimalla tai logaritmoimalla.

3.2 Selittävät muuttujat

Mahdollisina selittävinä muuttujina kokeilemme seuraavia:

- Velanhoitokulut ja kuluttajahintaindeksi
- 3 kk:n, 24 kk:n ja 120 kk:n korot
- Öljyn hinta ja bruttokansantuotteen muutos
- Standard & Poors -osakeindeksi

Nämä kaikki ovat myös Yhdysvalloista. Kuvassa 2 on selitettävien muuttujien havainnot. Bruttokansantuote ja velanhoitokulut olivat alun perin ilmoitettu 3 kuukauden välein. Ne on muutettu 1 kk välisiksi havainnoiksi laittamalla 3 samaa arvoa peräkkäin.



Kuva 2. Selitettävien muuttujien havainnot kuukausittain alkaen helmikuusta 1990 ja päättyen joulukuuhun 2006

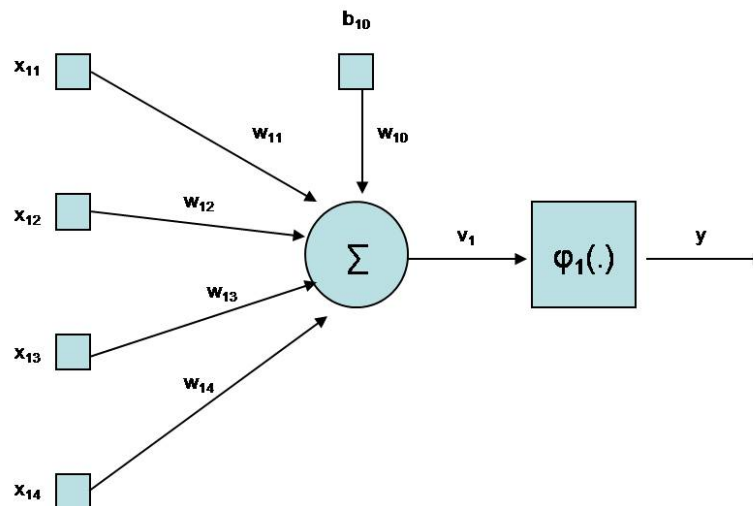
4 Kiinteistömarkkinoiden mallintaminen

Kirjallisuustutkimuksessa löytyneiden artikkelien perusteella valittiin kolme kehittävää mallia perustuen neuroverkkoihin, ARIMAX- sekä GARCH-malleihin.

4.1 Neuroverkko

Neuroverkon perusajatus on matkia aivojen datan käsittelymenetelmiä. Neuroverkko koostuu neuroneista ja niiden välisistä synapseista, biologisen esikuvan mukaan. Käytännössä neuroni on yksinkertainen matemaattinen funktio ja synapsi on painokerroin. Neuroverkoista voidaan tehdä hyvinkin monimutkaisia. Esimerkiksi synapsit voivat mennä ristiin, muodostaa takaisinkytkentöjä tai sisältää aikaviiveitä. Multi Layer Perceptron:ssa (MLP) neuronit ovat järjestetty tasoihin eivätkä synapsit sisällä viiveitä eivätkä mene ristiin. Tässä työssä käytetään vain MLP-neuroverkkoa.

Tarkastellaan nyt yksittäistä neuroniam, Kuva 3. Neuroniam tuleee 4 inputtia ja 1 bias-termi. Nämä kaikki kerrotaan painokertoimilla ja summataan yhteen. Tuloksena saadaan output v_1 . Tämä toimii aktivaatiofunktion $\varphi_1(\cdot)$ inputtina. Aktivaatiofunktion on yleensä epälineaarinen. Funktion ulostulona saadaan y , joka on siis neuronin ulostulo.



Kuva 3. Yksittäinen neuroni.

MLP:ssä neuronit ovat järjestetty tasoihin ja edellisen tason neuronien ulostulot toimivat seuraavan neuronitason inputteina.

Neuroverkon opetuksessa voidaan käyttää useita eri menetelmiä, tarkastellaan tässä vain backpropagationia, BA. BA algoritmi jakautuu kahteen vaiheeseen, ”forward passing” ja backward passing”. Forward-vaiheessa verkkoon syötetään inputit ja verkko muodostaa niiden perusteella vasteen. Backward-vaiheessa verkon vastetta verrataan haluttuun

referenssisignaaliin ja niiden erotuksen perusteella päivitetään painokertoimet. Tätä toistetaan kunnes etukäteen annetut tavoitteet toteutuvat. Näitä tavoitteita voivat olla tarpeeksi pienen signaalien erotus tai kunnes verkkoa on opetettu tarpeeksi kauan.

Neuroverkoilla on monia hyviä puolia verrattuna perinteisiin tilastollisiin menetelmiin. Esimerkiksi käsiteltävä ongelma ei tarvitse olla hyvin strukturoitu, havaintoja voi puuttua, havainnot voivat olla epälineaarisesti riippuvia keskenään ja havainnot ei tarvitse olla peräisin jostain jakaumasta. Ongelmaksi MLP:n kohdalla tulee se, kuinka monta piilotettua kerrosta verkossa on ja kuinka monta neuronit käytetään kussakin kerroksessa. Mikäli neuronien määrä on liian korkea, neuroverkko ylioppii aineiston eikä pysty yleistämään oppimaansa ennusteisiin. Toisaalta, jos neuroneja on liian vähän, ei opetus pysty löytämään riippuvuussuhteita muuttujien välillä. Toinen merkittävä ongelma neuroverkkojen kohdalla on se, että kyseessä on käyttäjän kannalta musta laatikko, eli painokertoimilla ei ole taloudellista tulkintaa. Käyttäjä ei siis tiedä mitä malli oikeastaan tekee ja vielä tärkeämpänä kysymyksenä miksi. Mallin oikeellisuuden ja luotettavuuden kannalta nämä ovat hankalia ominaisuuksia.

4.1.1 Tulokset

Käytetty data jaetaan ensin kahteen osaan, opetus- ja testidataan. Yleisenä sääntönä opetuksessa on ollut se, että mallin pitäisi ennustaa hyvin. Neuroverkko voidaan nimittäin opettaa siten, että opetusdatalla virhe on mahdollisimman pieni, mutta ennustamisominaisuudet ovat huonot. Koska EPRA:n indeksissä, öljyn hinnassa ja osakkeen indeksissä on eksponentiaalista kasvua, näistä muuttujista on otettu logaritmi. EPRA-indeksin kohdalla ennen logaritmoitua on tehty seuraava operaatio:

$$y_t = \frac{x_t}{x_{t-1}}$$

Mitään muita transformaatioita ei tehty. Bruttokansantuotteesta ja velanhoitokuluista on saatu kuukausittaista asettamalla kolme samaa arvoa peräkkäin. Alkuperäinen datahan oli näiden osalta 3 kk välein. Neuroverkkojen osalta tällä muutoksella ei ole haittavaikutusta toisin kuin lineaarisissa regressiomalleissa.

Koska neuroverkon säädettäviä parametreja on tässä tapauksessa paljon, ei kaikkia kombinaatioita voida käydä läpi, joten on mahdollista löytää vielä parempia tuloksia. Näitä säädettäviä parametreja ovat esimerkiksi selitettävät muuttujat – mitkä niistä otetaan malliin mukaan ja kuinka monta edellistä havaintoa kyseisestä muuttujasta – kuinka monta tasoa on neuroverkossa ja kuinka monta neuronin per taso, mitä opetusalgoritmia käytetään, mitä funktioita käytetään aktivaatiofunktiona ja kuinka paljon dataa käytetään opetukseen. Tarkastellaan tässä muutamia hyväksi havaittuja kombinaatioita.

Neuroverkkomallin toteutus on tehty Matlabin Neural Network Toolboxin avulla. Käydään tässä läpi kaikkien mallien yhteiset ominaisuudet. Jokaisessa mallissa opetusalgoritmina on käytetty Scaled conjugate gradient backpropagation algoritmia. Ulostulotasossa neuroneja on vain yksi ja sen aktivaatiofunktio on lineaarinen siirtofunktio. Muiden neuronien aktivaatiofunktio on hyperbolinen tangenttifunktio.

4.1.1.1 Selitettävää selitetään omilla havainnoilla

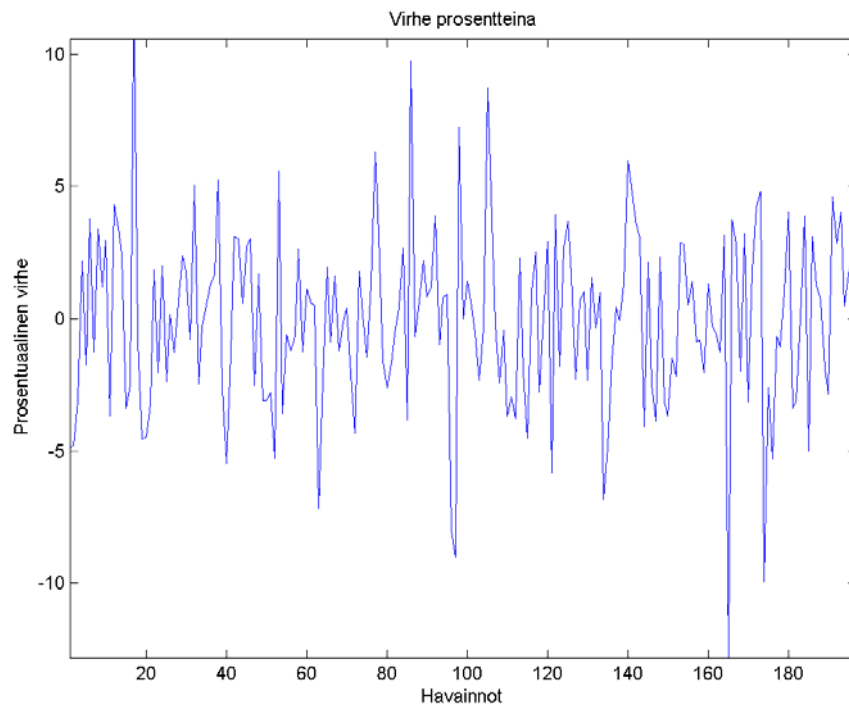
Tarkastellaan ensin tilannetta, jossa selitettävää muuttujaa yritetään selittää omien havaintojen avulla. Tässä tapauksessa edellä mainittua muunnosta EPRA-indeksille ei ole tehty. Opetusdatan osuus on 180 havaintoa ja testidata sisältää loput havainnot. Hyväksi malliksi osoittautui neuroverkko, jossa on yksi kahden neuronin piilokerros. Lisäksi ulostulotasossa on 1 neuronin, eli yhteensä neuroneja on 3 kappaletta. Edellisiä havaintoja syötetään vain yksi. Kuvassa 4 saadulla mallilla ennustettu havaintoja ja niitä on verrattu toteutuneisiin arvoihin. Kuvasta nähdään, että virhe muutamaa poikkeusta lukuun ottamatta pysyy 10 %:n sisällä. Kuvassa 5 on ennustetut havainnot ja toteutuneet arvot. Kuvassa on havaittavissa pientä viivettä ennusteessa, muuten malli toimii hyvin. Esitetään vielä opetuksen tuloksena saatu malli funktiona

$$y = e^{2.961 \cdot \tanh(0.274 \cdot x - 2.192)} + (-1.819) \cdot \tanh(-0.338 \cdot x + 1.319) + 6.385 \quad (1)$$

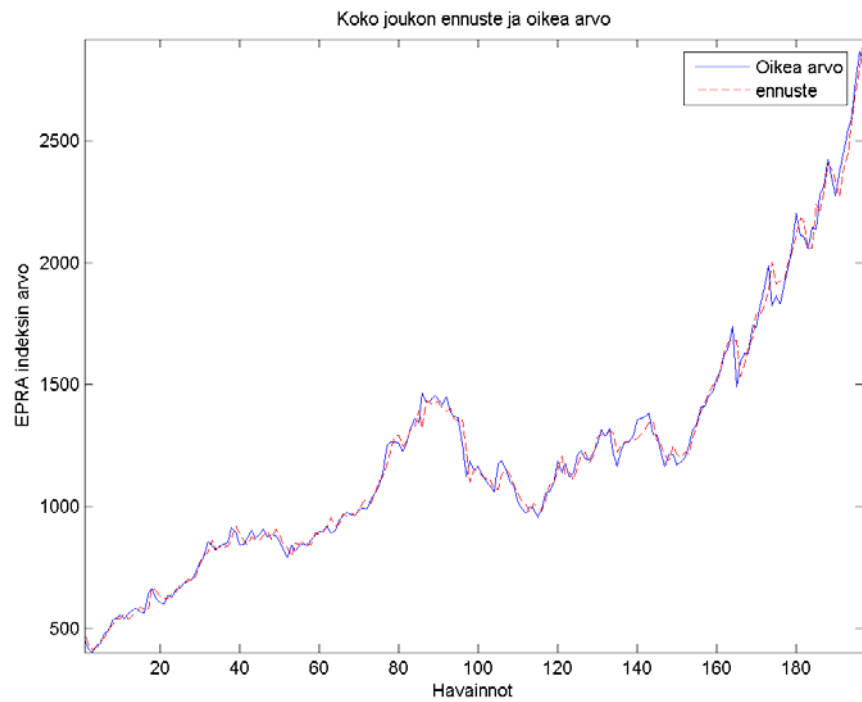
Koska itse ennustaminen ei ole mallinnuksessa useinkaan se tärkein tehtävä, vaan arvон muutoksen ennustaminen. Tästä eteenpäin käytetään EPRA-indeksin osalta edellä kuvattua muunnosta.

4.1.1.2 Selittäjinä; EPRA-indeksi, öljyn hinta, osake-indeksi ja bruttokansantuote

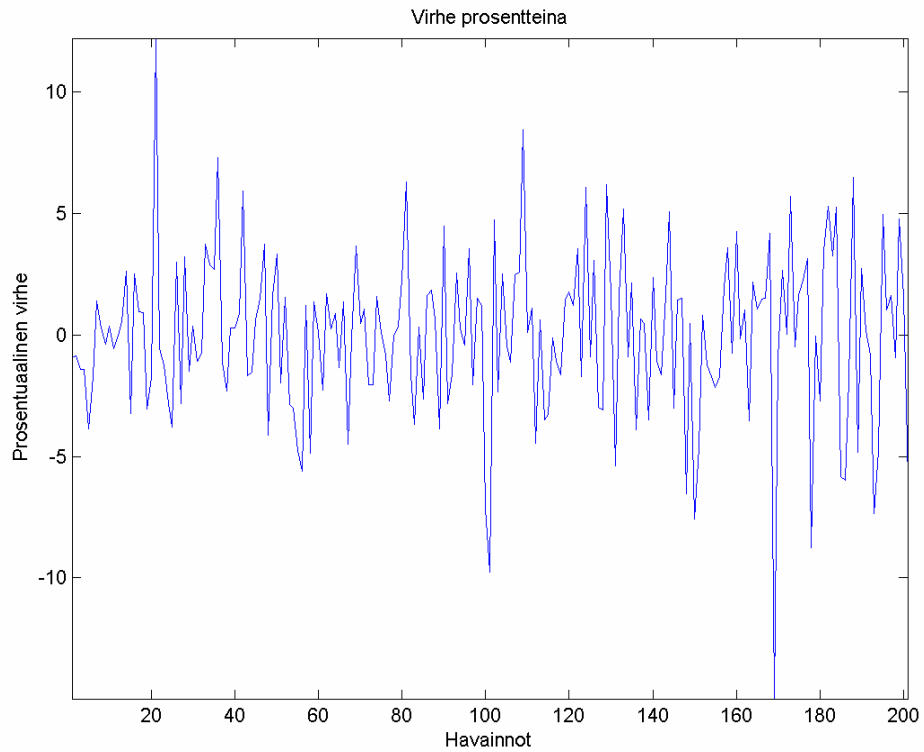
Seuraavaksi tarkastellaan tilannetta, jossa syötteenä on EPRA-indeksi, öljyn hinta, osake-indeksi ja bruttokansantuote. Kaikkia näitä muuttujia otetaan 2 edellistä. Ensimmäiseen piilokerrokseen valittiin 5 neuronua ja toiseen piilokerrokseen 4. Opetusdatan osuus on 180 havaintoa ja testidata sisältää loput havainnot. Kuvassa 6 on ennusteen virhe ja kuvassa 7 on ennustetut toteutuneet arvot. Lopputuloksena saadun mallin funktiota ei tässä esitetä, koska se on muuten sama kuin edellä mutta nyt eksponenttissa useita samantyyllisiä lausekkeita kuin edellä. Selitysaste on tällä datalla 31.25 %.



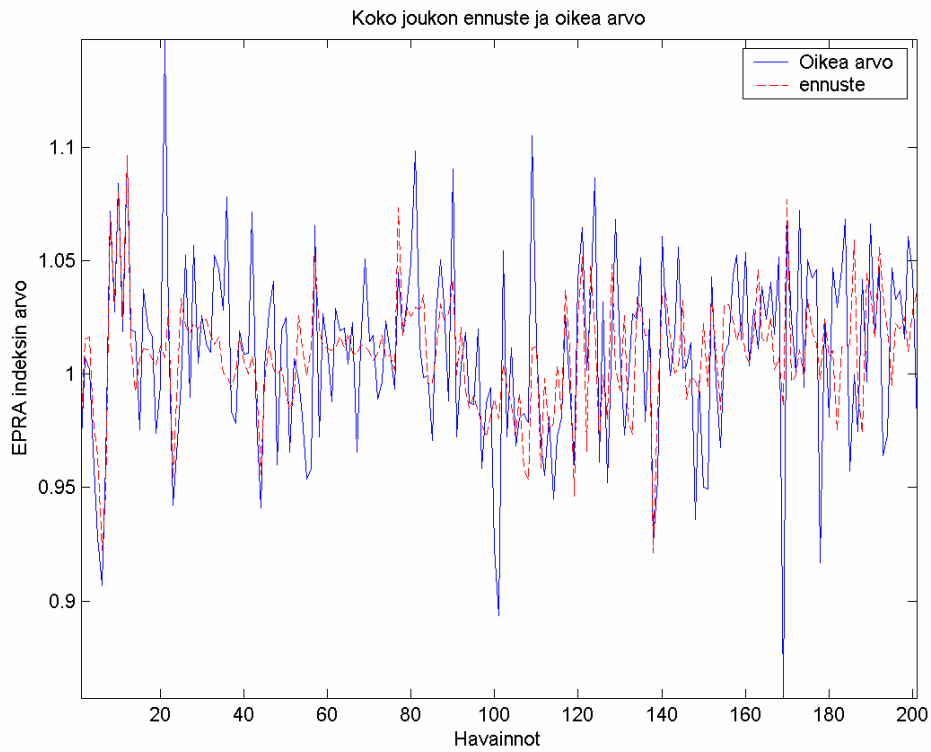
Kuva 4 Ennustettujen arvojen virheet prosentteina kuukausittain (helmikuu 1990 – joulukuu 2006)



Kuva 5 Ennustetut ja toteutuneet arvot (helmikuu 1990 – joulukuu 2006).



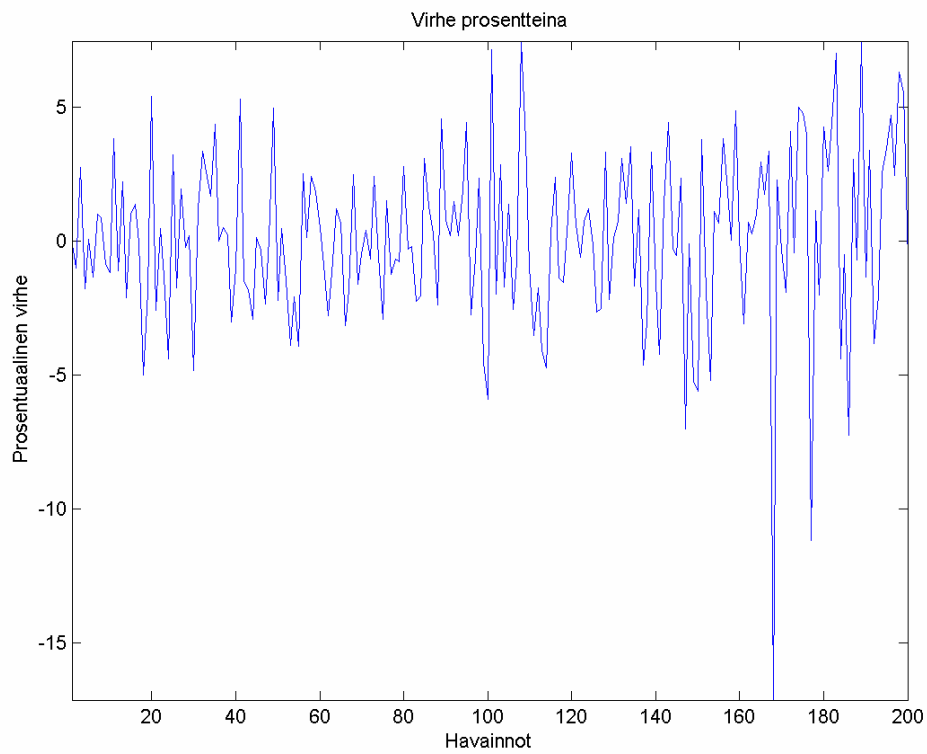
Kuva 6 Ennustettujen arvojen virheet prosentteina kuukausittain (helmikuu 1990 – joulukuu 2006) .



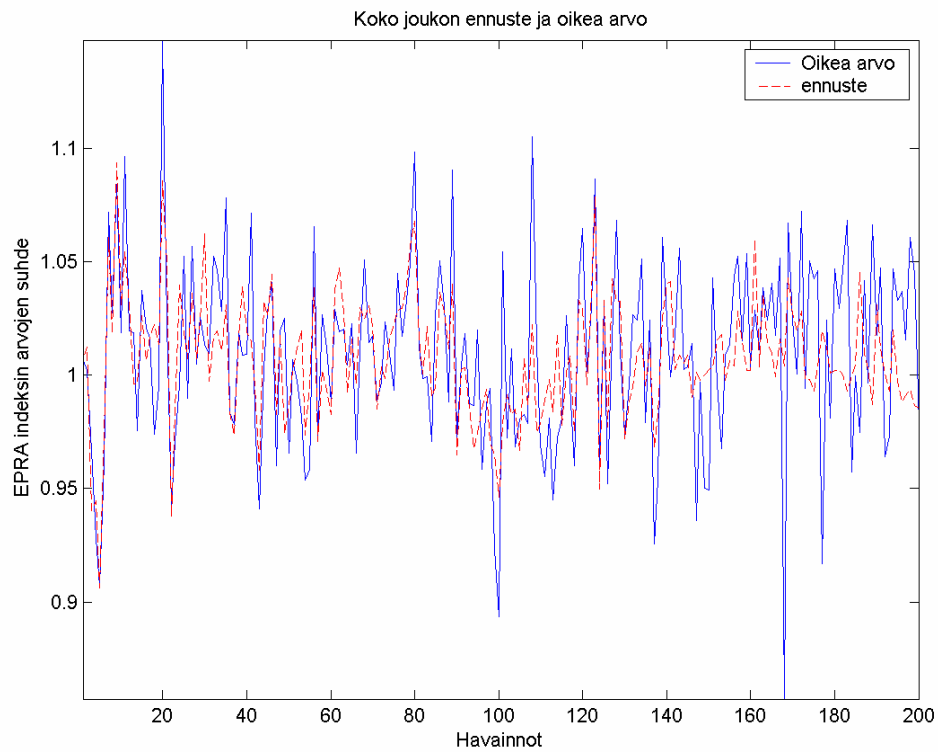
Kuva 7 Ennustetut ja toteutuneet arvot kuukausittain (helmikuu 1990 – joulukuu 2006)

4.1.1.3 Kaikki muuttujat mukana

Tehdään lopuksi mallinnus, johon otetaan kaikki selittäjät mukaan ja kaikista muuttujista 3 edellistä havaintoa. Neuroneja on 4 ensimmäisessä 3 toisessa tasossa. Tulokset ovat kuvissa 8 ja 4005. Selitysasteeksi saadaan noin 39.79 %.



Kuva 8 Ennustettujen arvojen virheet prosentteina kuukausittain (helmikuu 1990 – joulukuu 2006)



Kuva 9 Ennustetut ja toteutuneet arvot (helmikuu 1990 – joulukuu 2006).

4.2 ARIMAX

Eräs yleisimmistä tavoista mallintaa kiinteistömarkkinoiden käyttäytymistä makroekonomisten muuttujien avulla on ollut dynaamisten regressiomallien käyttö. ARIMAX (integrated autoregressive moving average model with exogenous variables) on osoittautunut toimivaksi malliksi erityisesti historiadatan selittämisessä, vaikkakaan sen ennustamisominaisuudet eivät ole parhaat mahdolliset. Kyseessä on dynaamisen regressiomallin (eli viivästettyjä selittäviä muuttujia sisältävän mallin) ja ARMA-mallin (autoregressiivinen liukuvan keskiarvon malli) yhdistelmä. Malli sisältää siis valitut selittävät muuttujat tiettyyn viiveeseen asti sekä vastaavasti tietyn asteisen autoregression ja virhetermien liukuvan keskiarvon osuuden. MA-komponentti kuvaa käytännössä aiempien satunnaisten kysynnän tai tarjonnan shokkien tai virheiden vaikutuksia ja AR-komponentti aikasarjan historian vaikutusta.

ARIMAX-mallissa käytetyn aikasarjadatan tulee olla stationaarista (jotta se voisi olla jonkin ARIMAX-prosessin generoimaa). Stationaarisuuteen vaaditaan kolme tekijää:

1. aikasarjan odotusarvo on vakio ajassa
2. aikasarjan varianssi on vakio ajassa
3. kovarianssi kahden hetken välillä riippuu vain ajanhetkien väliajasta, ei itse ajanhetkistä

Aikasarjassa ei saa siis näkyä trendiä, kausivaihtelua tai varianssin systemaattista vaihtelua. Tämä asettaa samat vaatimukset kuin tavallisille ARMA-malleille: AR-viivepolynomin juurien on oltava yksikköympyrän ulkopuolella. Käytännössä stationaarisuutta tutkitaan aikasarjan autokorrelaatio- ja osittaisautokorrelaatiofunktioiden avulla. Mikäli selitettävä aikasarja tai jokin selittäjistä on epästationaarinen, se on erikseen stationarisoitava differensoimalla.

$$\Delta Y_t = a + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \dots + \beta_p \Delta Y_{t-p} + \sum_{j=1}^m \left(\sum_{i=0}^{K_j} \gamma_{j,i} X_{j,t-i} \right) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Kaavassa 3, Δ on ensimmäinen differenssioperaattori ja kaavan ensimmäinen osa kuvaa AR-osaa, jossa Y on selitettävä muuttuja, jonka suurin viive on p . $X_{j,t-i}$ on puolestaan ulkoiset

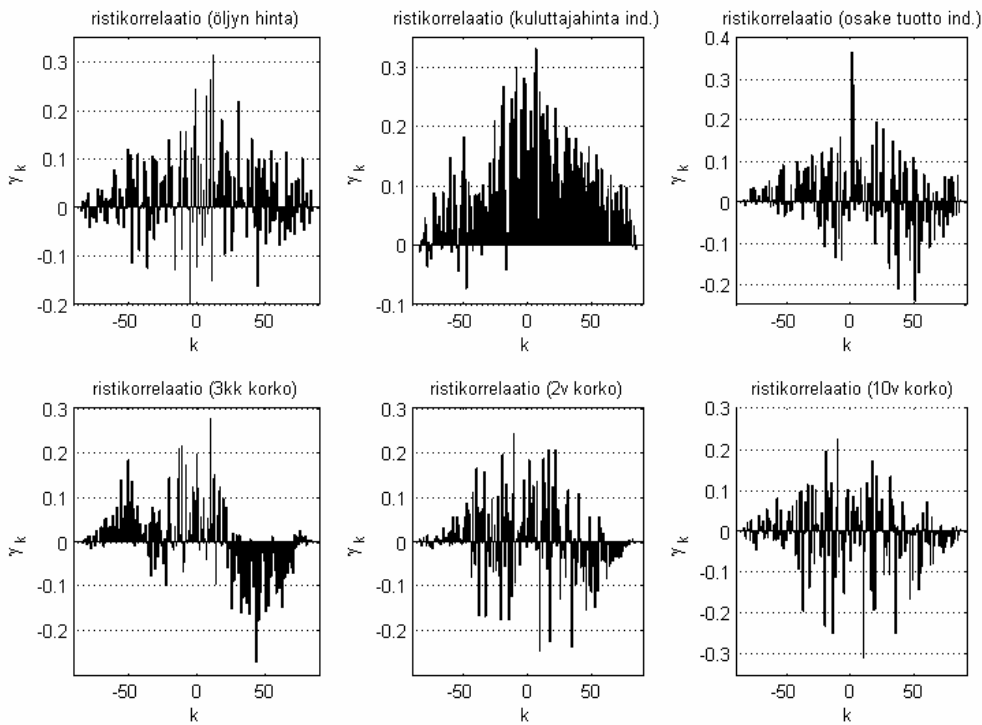
selittävät muuttujat sisältävä vektori. Näitä selittäjiä koko on yhteensä m kappaletta ja kullekin on määritelty oma maksimiviive K_j . ε_{t-q} kuvaa mallin MA-osaa, jonka suurin viive q .

Nyt selitettävässä indeksissä on havaittavissa selvä trendi, joten se on differensoitava. Auto- ja osittaisautokorrelaatiofunktioita tutkimalla havaitaan, että kyseessä on AR(1)-prosessi. Myös selittävissä muuttujissa havaitaan selviä trendejä, joten nekin on differensoitava. Ensimmäiset differenssit ovat riittäviä.

4.2.1 Tulokset

Ennen kuin ARIMAX-mallia voidaan varsinaisesti lähteä sovittamaan käyttämäämme dataan, on dataa hiukan esikäsiteltävä. Aluksi paras analyysitapa on katsella dataa paljain silmin ja etsiä siitä mahdollisia toistuvia säännönmukaisuuksia ja trendejä. Kiinteistöjen hintakäyrässä (kuva 1) on selkeästi havaittavissa eksponentiaalinen trendi. Tämän poistaminen tapahtuu datan logaritmoinnilla. Lisäksi tärkeätä ARIMAX-mallin kannalta on, että käytettävä aikasarja on stationarisoitu. Saadaksemme aikasarjasta stationaarisen on siitä järkevää laskea tuottoluvut ts. miten paljon asunnot ovat tuottaneet viimeisen ajan jakson aikana. Tämä tarkoittaa havainnon jakamista edellisellä havainnolla, ja saadun suhteen logaritmoimista luonnollisella logaritmilla.

$Y(i)$:n arvot kertovat nyt logaritmin datan muutoksesta edelliseen havaintoon verrattuna. Sekä selitettävä aikasarja, että selittävät aikasarjat päätettiin esittää tässä muodossa, sillä niiden muutokset muodostivat stationaarisen prosessin. Sopivaa mallia lähdettiin tässä tapauksessa etsimään käyttäen ylöspäin askellusta. Tämä tarkoittaa, että aluksi mukaan otetaan vain vähän selittäjiä ja parametreja, ja vähitellen määrää lisätään yrittäen koko ajan parantaa mallin selitystasetta. Ennen mallien sovittamista datalle tehtiin kuitenkin ristikorrelaatio-analyysi, sopivien viiveiden löytämiseksi. Ristikorrelaatiot laskettiin selitettävän ja jokaisen selittäjän välille. Ristikorrelaatioista saatiin kuvan 10 mukaiset.



Kuva 10: Ristikorraatiot selitettävän $Y(t)$:n ja selittäjien $X_j(t)$ välillä. Vaaka-akselin k tarkoittaa viivettä kuukausina.

Lyhyen tarkastelu jälkeen voidaan havaita ristikorraatioissa suurimpien piikkien osuvan taulukon 1 mukaisiin viiveisiin.

Taulukko 1: Selittäjien suurimmat ristikorraatiot

	viive (kk)	Korr. arvo
öljyn hinta	12	0,31
kuluttaja ind.	6	0,33
osaketuotto	0	0,36
3kk korko	10	0,27
2v korko	16	0,20
10v korko	9	0,30

Ensimmäisessä vaiheessa yritettiin sovittaa dataan pelkkää ARX-mallia, siten että selitykseen käytettiin jokaista muuttujaa yksitellen. Parasta selitystasetta haettiin muuttelemalla AR- ja ulkoisten selittäjien parametrien lukumääriä p ja K_i hiukan. Analyysin tulokset on nähtävissä taulukossa 2.

Taulukko 2: Yhden selittäjän ARX-mallien kertoimet

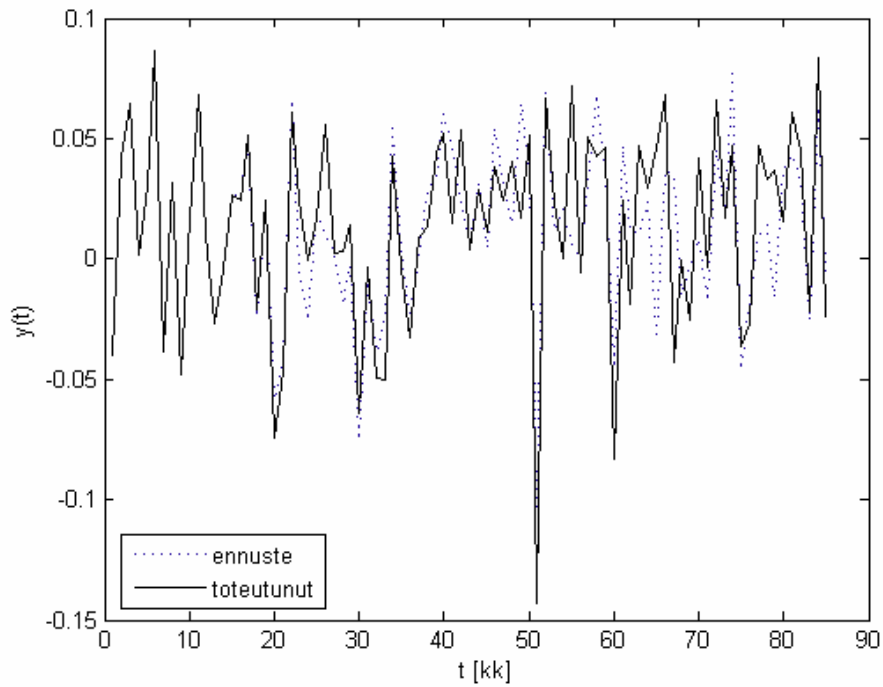
	p	K_i	i	selitys-%
öljyn hinta	3	3	10	18
kuluttaja ind.	3	4	6	11
osaketuotto	3	3	0	13
3kk korko	3	3	10	13
2v korko	3	3	16	16
10v korko	3	4	9	13

Ensimmäisen vaiheen analyysien jälkeen yhdistettiin saadut tulokset yhdeksi kaikki selittäjät sisältäväksi malliksi. Parametrien lukumäärien kohdalla tehtiin enää pieniä muutoksia. Parhaaksi lopputulokseksi tuli lopulta taulukon 3 mukainen malli, missä Siis AR-osa on kolmatta astetta ja selittäjät astetta 2-4.

Taulukko 3: ARX-mallin parametrien arvot ja eri selittäjille käytetyt viiveet

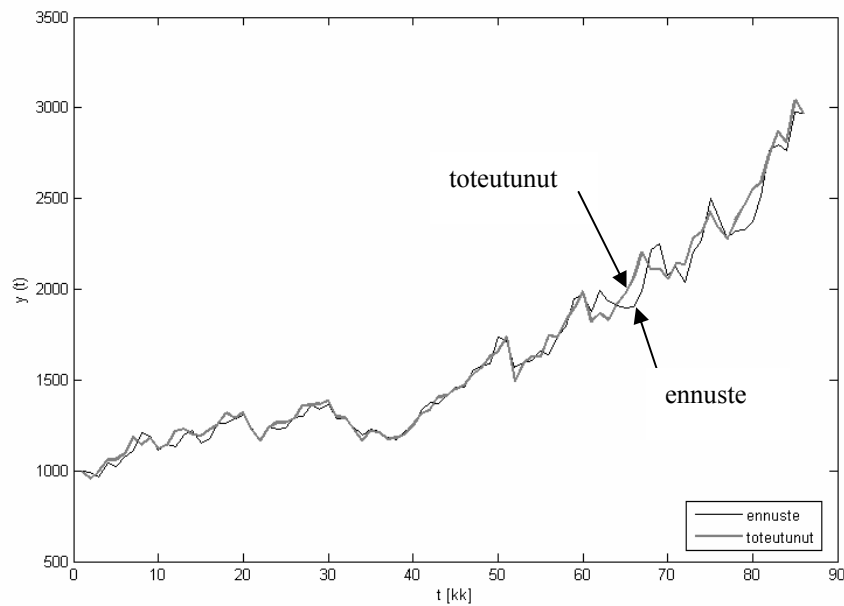
AR-osa		öljyn hinta		kulut. hinta indeksi		osake indeksi		3kk korko	2v korko		10v korko		
viive (kk):		10		6		0		10	16		9		
β_1 :	-0.1015	$\gamma_{1,1}$:	0.08988	$\gamma_{2,1}$:	5.211	$\gamma_{3,1}$:	0.2024	$\gamma_{4,1}$:	0.06096	$\gamma_{5,1}$:	0.07027	$\gamma_{6,1}$:	-0.1145
β_2 :	0.2367	$\gamma_{1,2}$:	-0.0704	$\gamma_{2,2}$:	3.859	$\gamma_{3,2}$:	0.5503	$\gamma_{4,2}$:	-0.1467	$\gamma_{5,2}$:	-0.04877	$\gamma_{6,2}$:	0.01071
β_3 :	0.1824	$\gamma_{1,3}$:	0.06608			$\gamma_{3,3}$:	0.07862	$\gamma_{4,3}$:	0.00374			$\gamma_{6,3}$:	0.03314
												$\gamma_{6,4}$:	0.1243

Saadulla mallilla suoritettu simulaatio näyttää kuvan 11 mukaiselta. ARIMAX -malli on laskettu kuvan 60 ensimmäisellä havainnolla ja loput 26 kk arvot on ennustettu tällä mallilla. Mallin selitysasteeksi saatiin 43,5 %.



Kuva 11 ARX-mallin ennuste ja toteutunut indeksin arvo kuukausittain tammikuu 2000 – tammikuu 2007

Mallin avulla piirretty Kiinteistöjen hintakäyrä on kuvassa 12. Malli osaa siis ennustaa toteutunutta hintaindeksiä melko hyvin.



Kuva 12: ARX-mallin antama ennuste ja toteutunut hintakäyrä kuukausittain tammikuu 2000 - tammikuu 2007

4.3 Garch-malli

Autoregressive Conditional Heteroskedasticity, eli ARCH, on aikasarjamalli, jossa havaintojen volatilitteetti muuttuu ajan funktiona [Pin98]. Esimerkiksi seuraavanlaiselle regressioyhtälölle

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t \quad (4)$$

termin ε_t varianssi voi riippua termistä X_{2t}^2 . Yleisemmin varianssi voi riippua vaikka kuinka monesta havaintoarvosta ja niiden volatilitteeteista eri periodeissa. Yleistettynä p:n eri periodin volatilitteeteista riippuvalle ARCH(p) mallille voidaan volatilitteetti kirjoittaa seuraavassa muodossa

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \quad (5)$$

jossa α_0 on vakio ja $\alpha_n \varepsilon_{t-n}^2$ on hetken n volatilitteetista riippuva tekijä. Mallissa ε_t^2 :t ovat heteroskedastisia ehdoilla ε_{t-n}^2 . Molemmat yhtälöt (4) ja (5) voidaan ratkaista suurimman uskottavuuden menetelmällä.

Jos oletetaan, että yhtälö (4) on vain jakautunut viivemalli σ_t^2 :stä, niin monet ε_t^2 :t voidaan korvata vain muutamalla σ_t^2 :llä. Tämä yksinkertaistettu versio on niin sanottu Generalized Autoregressive Conditional Heteroskedasticity, eli GARCH, malli.

GARCH-malli toimii parhaiten melko stabiilissa markkinatilanteessa, vaikka se ottaakin huomioon volatilitteetin muutokset [Mat]. Malli ei usein toimi hyvin epänormaaleissa tilanteissa ja niiden aikana se yleensä ennustaa liian huonosti ääritilanteita. GARCH malli edellyttää riittävästi tietoa ja on melko herkkä datan määrän vaihteluille pienillä havaintojen lukumäärillä.

4.3.1 Tulokset

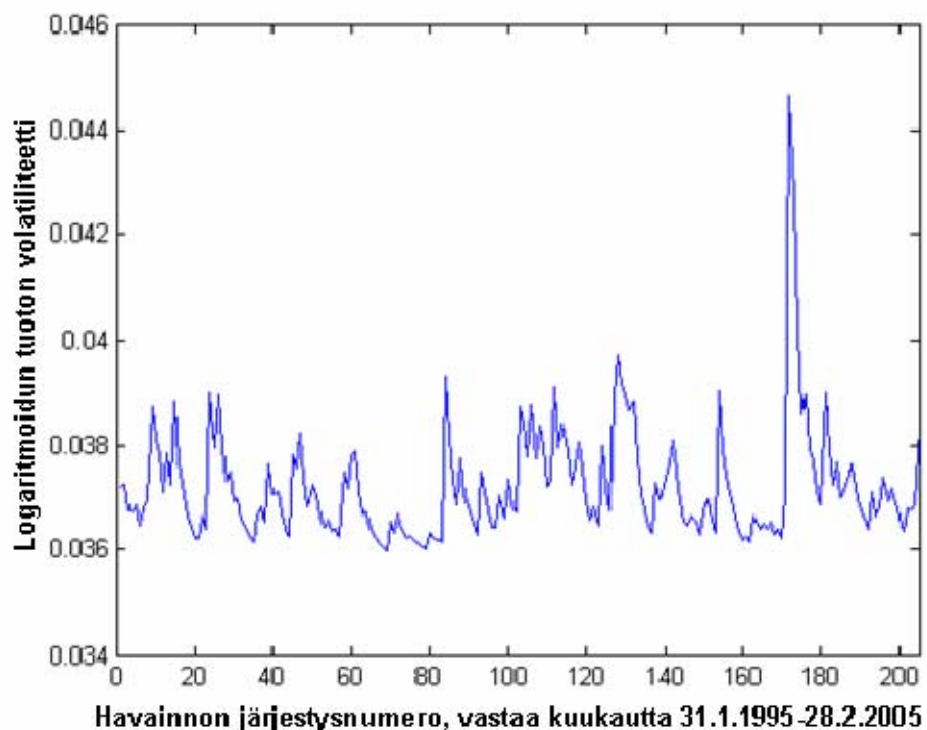
4.3.1.1 GARCH-mallin muodostus

GARCH-malli muodostettiin käyttäen Matlabin valmiita GARCH-työkaluja. Koska malli on hyvin herkkä havaintoarvojen lukumäärälle, jouduimme jättämään pois havaintoarvot, joita oli saatavilla vain kolmen kuukauden välein. Aikasarjaa meillä oli saatavilla kuukausittain 31.1.1990 alkaen, mutta koska 90-luvun alussa oli jonkinasteinen taantumus, niin jätimme

osan datasta pois (ajankohtaan 31.1.1990–31.12.1994 viitataan tekstissä monesti termillä ”lamavuodet”). Sen sijaan käytimme havaintoja vuosilta 31.1.1995–28.2.2007. Mallin selittävinä arvoina on yleensä käytetty havaintoja 31.1.1995–28.2.2005. Mallin avulla on tehty ennuste ajalle 31.3.2005–28.2.2007.

Mallin selitettävänä tekijänä siis on EPRA:n NAREIT-indeksi. Se on Pohjois-Amerikan kiinteistöihin sijoittavien rahastojen indeksi. Selitettävänä tekijöinä kokeillaan öljyn hintaa (USD), kuluttajahintaindeksiä, Standard & Poors osakeindeksiä (S&P 500), 3-kuukauden korko, 2 vuoden korko ja 10 vuoden korko.

Kaikista muuttujista lasketaan kahden periodin välinen muutos ja saatu arvo logaritmoidaan (luonnollinen logaritmi), jotta saadaan paremmin käyttäytyvä malli ja jotta saadaan paremmin selville kuinka hyvin malli oikeasti havaitsee pienet muutokset. Kuvassa 13 on esitelty kiinteistöindeksin volatilitteetti ajan funktiona. Kuva osoittaa selkeästi, miksi GARCH-malli voisi toimia hyvin ongelman mallintamisessa.



Kuva 13 Kiinteistöindeksin tuoton logaritmin volatilitteetti

Itse malli luodaan Matlabilla, joka muodostaa alkuarvojen perusteella mallin sekä keskiarvoille, että variansseille. Mallin avulla teemme 24 kuukauden ennusteen kiinteistöindeksille. Ennustetta varten luodaan aikaisempien havaintojen avulla jakauma residuaaleille. Ennuste käyttää hyväkseen tämän residuaalijakauman arvoja määrittäessään tulevaisuuden arvoja. Koska tämä johtaa lievään satunnaisuuteen, niin laskemme kerralla useita eri ennusteita (laskemme 24 eri ennustetta). Näiden useiden ennusteiden avulla pystymme tekemään ”keskiarvoennusteen” ja pystymme tarkastelemaan, minkä verran yksittäiset ennusteet poikkeavat toisistaan.

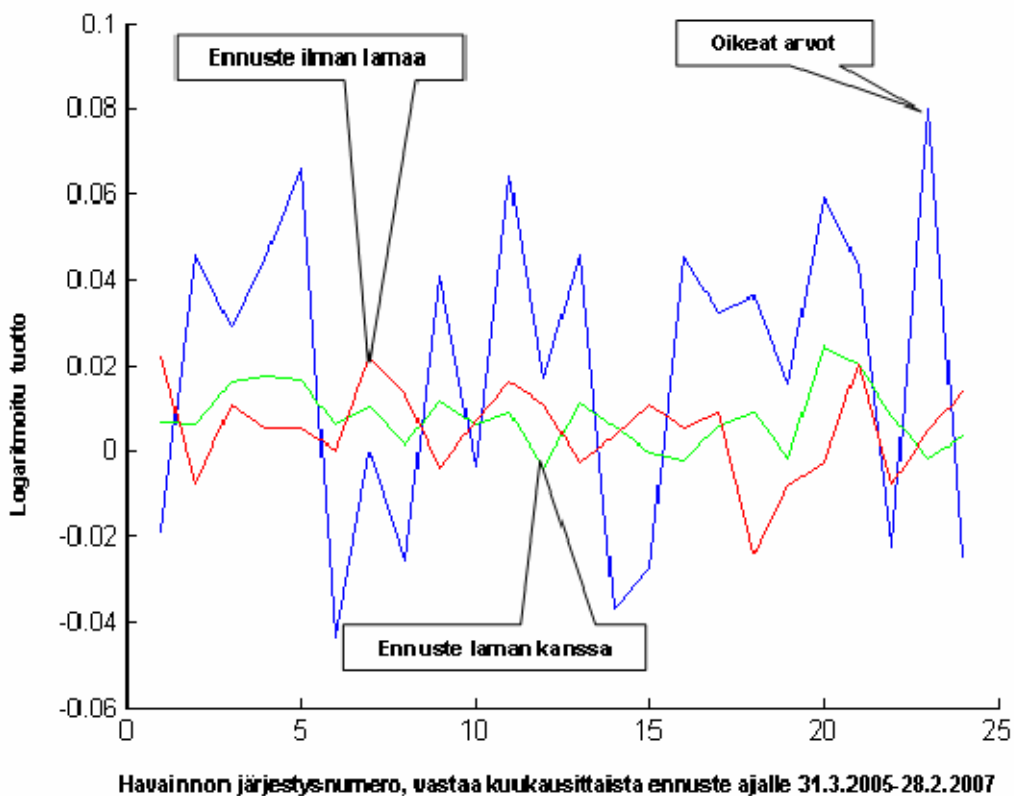
Laskemme selitysasteen ottamalla neliön 24 kuukauden ennusteen ja näitä 24 kuukautta vastaavien oikeiden arvojen välisestä korrelaatiosta. Ennustamisen kannalta korrelaatio on nimenomaan kiinnostava, sillä indeksin absoluuttista arvoa kiinnostavampaa (ja varmempaa) on tietää sen muutoksen trendi. On tärkeää myös huomata, että kaikki saadut lukuarvot perustuvat logaritmiselle tuottoasteelle, ei oikeille havaintojen arvoille.

4.3.1.2 GARCH-mallin ominaisuuksien esittely

Erilaisia malleja kokeilemalla ja etenkin erilaisia aikasarjoja kokeilemalla havaitsimme, että malli on hyvin herkkä aikasarjojen pienillekin muutoksille. Aikasarjojen pituuden ja ajankohdan valinnalla näyttää olevan suurempi merkitys mallin tuloksiin kuin esimerkiksi selittävien tekijöiden valinnalla. Tämän takia mallin tuloksia pitää tarkastella kvalitatiivisesti muun tiedon rinnalla, ja mallin käyttäjällä on hyvin suuri vastuu tulosten raportoinnissa. Erityisesti mallin rajoitteet näkyivät hyvin eri aikasarjojen ajankohtia kokeiltaessa, sillä jos 90-luvun alun lamavuodet otettiin mukaan aikasarjoihin, mallin selityskyky tippui huomattavasti. Ennusteissa on huomattava, että käyttämämme malli ei tee ennustetta käyttäen askeltavaa ennustusmenetelmää. Eli oikeita havaittuja arvoja ei käytetä selittämään seuraavan askeleen ennustetta. Tämä huonontaa ennusteiden oikeellisuutta, mutta tuo paremmin esille riskit, jotka liittyvät pitkän aikavälin ennusteeseen silloin kun oikeita havaintoja ei ole käytettävissä.

Koska tiesimme mallin rajoitukset ja lamavuosien olemassaolon, niin tarkastelimme ensin, mistä vuodesta alkaen aikasarjaa pitäisi ottaa. Nopeasti kokeilemalla havaitsimme, että 90-luvun ensimmäiset vuodet kannattaa tarkastella erikseen. Tässä on nimittäin huomattava, että

vaikka lamavuosien pois tiputtaminen pienentää havaintojen lukumäärää, se kuitenkin antaa volatilitteetille enemmän tilaa liikkua. Tämä johtuu siitä, että ennustetta tehdessä mallissa käytetään aikaisempia havaintoja selittämään mahdollisia tulevaisuuden arvoja. Kuvassa 14 on esimerkki, miten lamavuodet vaikuttavat ennustettuun malliin. Vihreä käyrä on ennuste laman kanssa, punainen ilman lamaa ja sininen on oikeat arvot

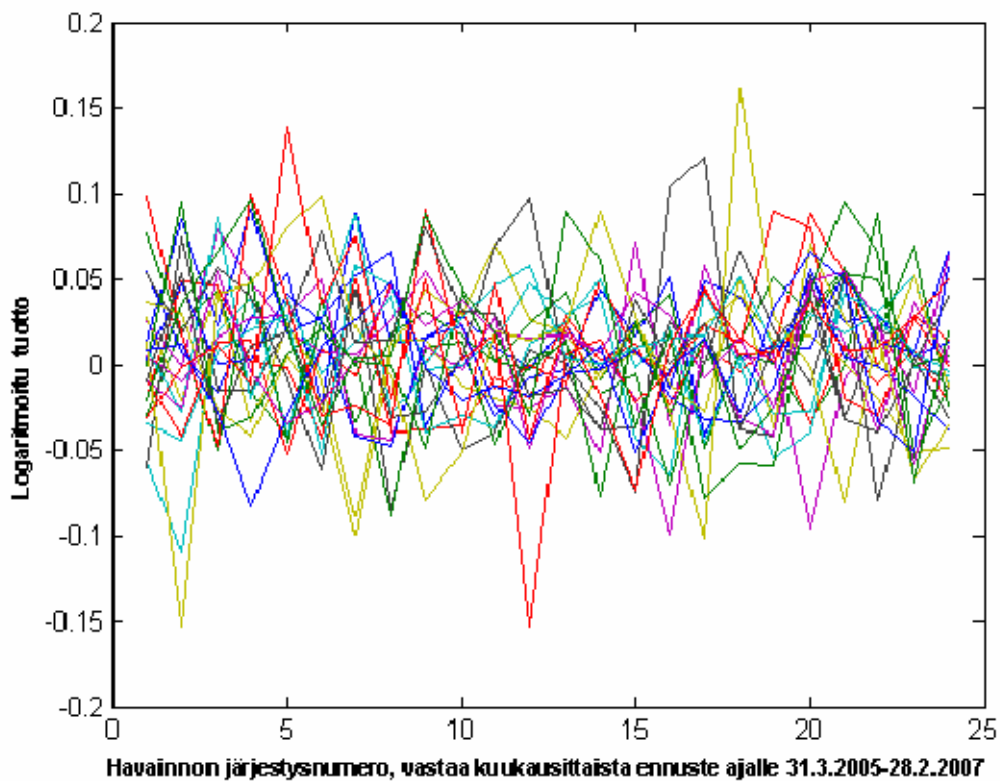


Kuva 14 Lamavuosien mukanaolon vaikutus ennusteeseen

Useissa malleissa ja eri aikasarjoja kokeilemalla havaitaan nopeasti, että parhaimmat selittäjät ovat S&P osakeindeksi ja 3 kuukauden korko. Valitut selittäjät vastaavat hyvin saatavilla olevaa a priori-tietoa. Lisäksi näiden selittäjien paremmuutta korostaa niiden ennustettavuus. Koska mallin tarkoitus on auttaa ennustamaan kiinteistöjen tuottoa useiden kymmenien vuosien päähän, ei olisikaan mitään järkeä käyttää selittäjänä muuttujaa, jota ei itsessään pystytä mitenkään ennustamaan (vrt. esim. öljyn hinta).

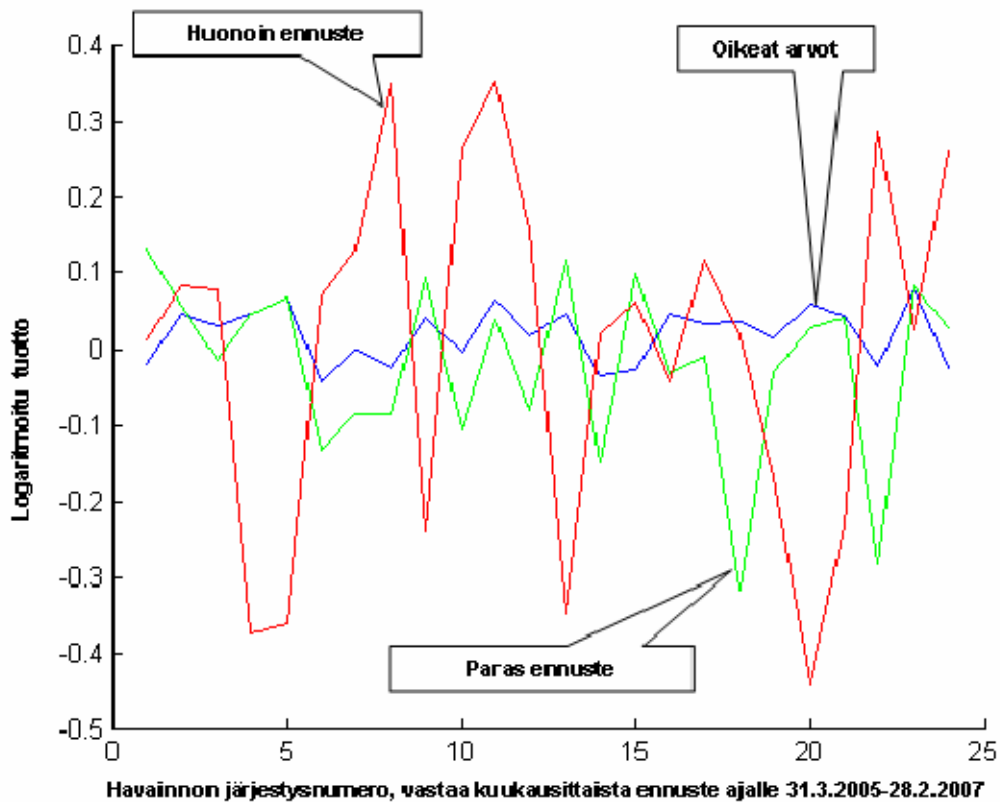
Kuvassa 15 on esimerkki eri ennusteiden poikkeamista. Tässä on siis 24 eri ennustetta aivan täsmälleen samoilla parametreilla (selittäjänä osakeindeksi). Vaikka kuva on jo vähän

epäselvä, niin se osoittaa, miten erilaisia tuloksia voidaan saada ja miten suuri merkitys on sillä, mikä näistä ennusteista otetaan kulloinkin käsittelyyn. Valitun ennusteen hyvyden me siis pystymme tarkistamaan vertaamalla oikeisiin arvoihin, mutta jos ennustetta käyttää puhtaasti ennusteena, niin silloin tulisi mielestämme tarkastella keskiarvoa. Tällä tavalla pystytään minimoimaan valittuun ennusteeseen liittyvää virhettä.



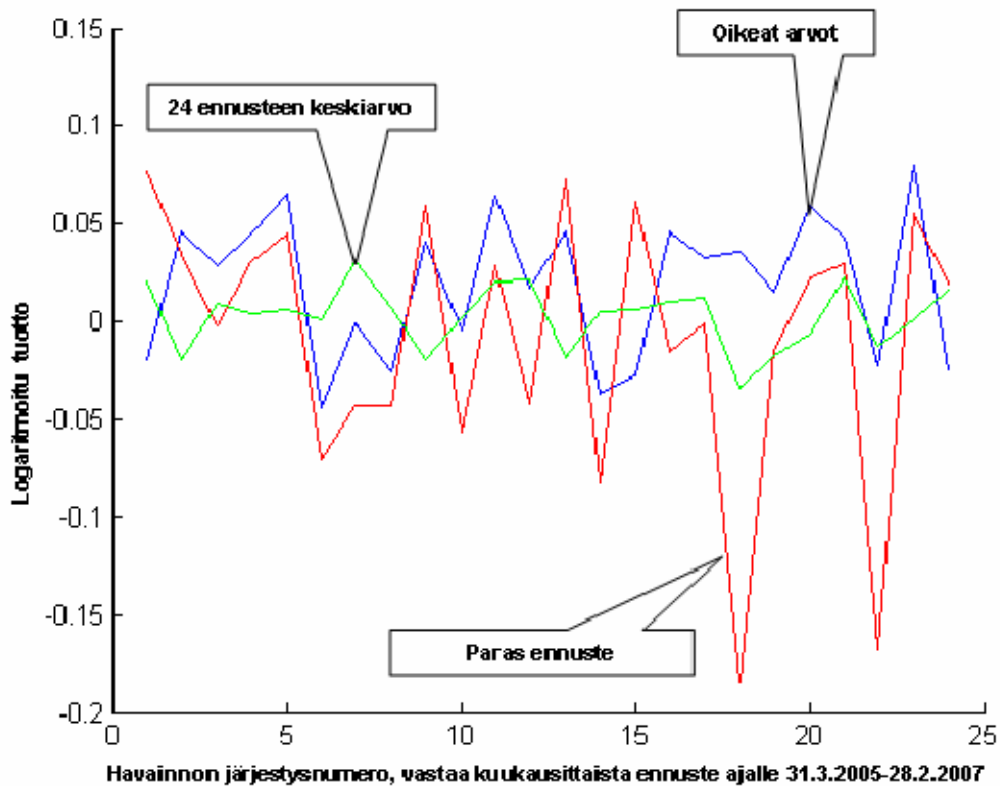
Kuva 15 24 ennusteen tulokset kun selittäjänä osakeindeksi

Kuvassa 16 on vielä selkeämmin esitetty edellä mainittu ongelma. Tämän kuvan parametreina ovat osakeindeksi ja kolmen kuukauden korko. Parametrit ovat eri näissä kuvissa, jotta osoittautuu selkeämmin kuinka systemaattinen ongelma tämä on ja jotta saamme paremmin osoitettua tämän ilmiön merkitys. Mukaan on otettu korrelaation (oikea osakeindeksi vs. ennuste) mukaan huonoin ja paras ennuste. Punainen on selkeästi huonoin ja siinä piikit ovat suurimpia ja monesti oikeaa aikasarjaa päinvastaisia (oikea kiinteistöindeksi on sininen). Vihreä on paras ennuste, joka noudattaa paikoitellen oikeita havaintoja melko hyvin, mutta myös tässä on havaittavissa selkeästi virheellisiä piikkejä.



Kuva 16 Paras ja huonoin ennuste kun selittäjänä osakeindeksi ja kolmen kuukauden korko

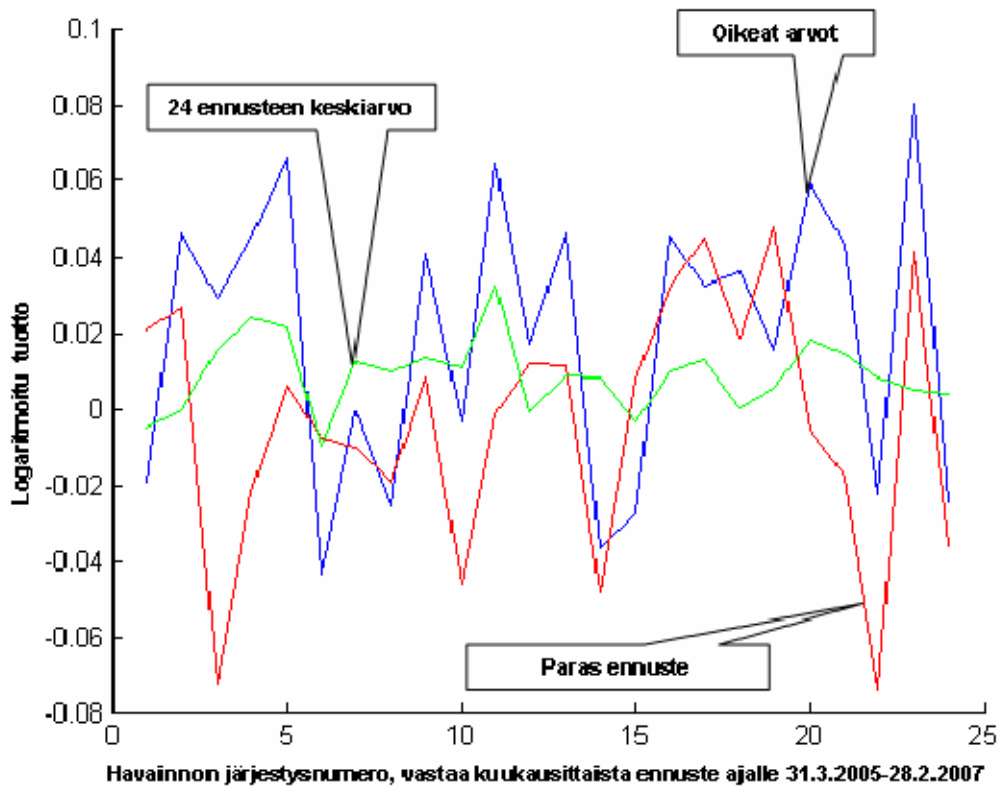
Kuvassa 17 on samat tulokset kuin kuvassa 16. Nyt on kuitenkin piirretty korrelaation mukaan paras ennuste (punainen väri) ja 24 ennusteen keskiarvo (vihreä väri). Oikeat kiinteistöindeksin arvot ovat sinisellä. Kuva kertoo selkeästi, kuinka keskiarvotus pienentää selvästi piikkejä. Kuitenkin on huomattavaa, miten keskiarvossa säilyy selkeästi pieniä piikkejä tietyissä kohtaa aikasarjaa. Tämä on merkittävää, sillä se osoittaa, että keskiarvon käyttäminen ei tarkoita suoran viivan sovittamista kuvaan.



Kuva 17 Paras yksittäinen ennuste ja 24 ennusteen keskiarvo kun selittäjänä osakeindeksi ja kolmen kuukauden korko

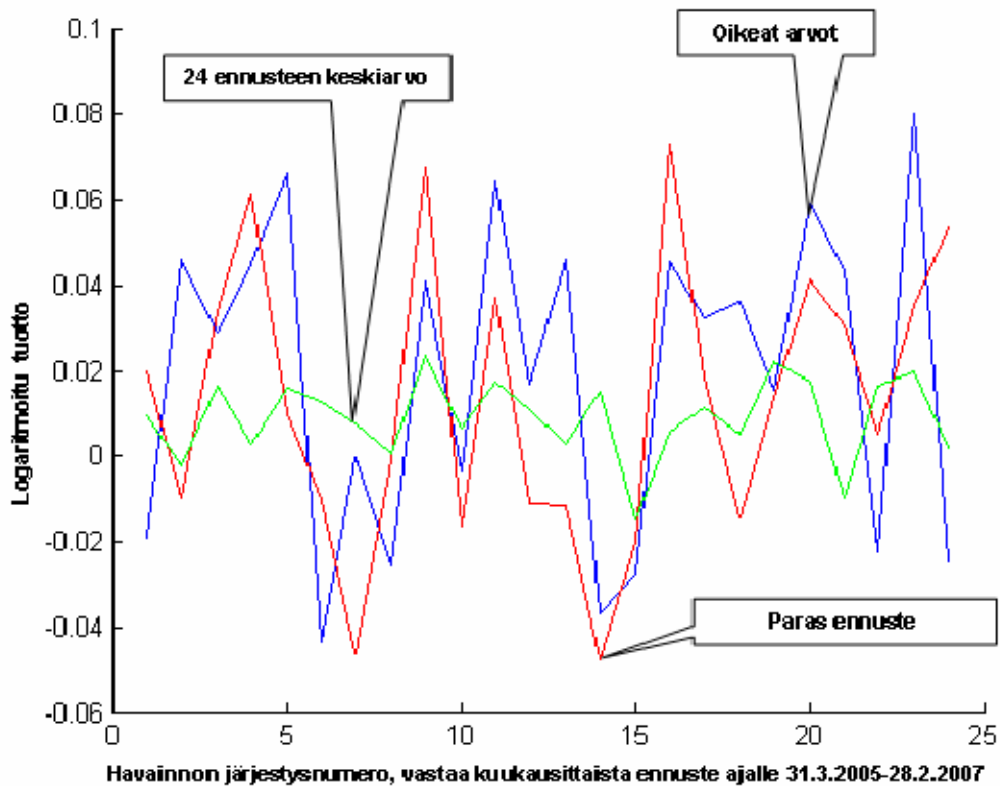
4.3.1.3 Valittu GARCH-malli ja sen tulokset

Selitysstetta tarkastelemalla löysimme parhaimmat selittäjät, jotka siis ovat osakeindeksi ja kolmen kuukauden korko. Liitteessä 2 on raportoitu tuloksia tälle mallille eri aikasarjoilla. Kuvassa 18 on selittävinä tekijöinä osakeindeksi ja kolmen kuukauden korkotaso. Lamavuodet on jätetty pois, mutta mukana mallin muodostuksessa on käytetty myös ennustamisaikavälin arvoja. Punainen väri kuvaa korrelaation mukaan parasta ennustetta (selitysstaste 0,19), vihreä 24 ennusteen keskiarvoa (selitysstaste 0,31) ja sininen oikeita indeksin arvoja. Paras ennuste antaa selkeästi keskiarvoa parempia ennusteita piikkien kohdilla, mutta toisaalta se antaa myös muutamia aivan väärän suuruisia tuloksia. Keskiarvo ei sen sijaan yhtä hyvin pysty ennustamaan piikkien suuruutta, mutta osuu paremmin kohdalleen piikkien paikkojen määrittämisessä.



Kuva 18 Paras ennuste ja ennusteiden keskiarvo osakeindeksin ja 3kk koron avulla ilman lamavuosia

Kuvassa 19 on tehty ennuste aivan samojen aikasarjojen mukaisella mallilla, mutta nyt mukana on lamavuodet. Jälleen kerran punainen on korrelaation mukaan paras ennuste (selitysaste 0,23), vihreä on 24 ennusteen keskiarvo (selitysaste 0,04) ja sininen on indeksin todelliset arvot. Edelliseen verrattuna huomataan, että paras malli mukautuu paljon paremmin piikkien kohdalla ja osaa ennustaa niiden suuruuden paljon paremmin. Keskiarvon ennuste ei ole kuitenkaan enää aivan yhtä hyvä kuin mitä se oli ilman lamavuosien mukaan ottamista.



Kuva 19 Paras ennuste ja ennusteiden keskiarvo osakeindeksin ja 3kk koron avulla lamavuosien kanssa

On kuitenkin huomattava, että edellä olevissa havainnoissa on siis mallin muodostamisessa mukana myös ennustettavan aikavälin arvot. Jos näitä arvoja ei ole mukana, niin ilman lamavuotta parhaan mallin selitysaste on 0,16 ja keskiarvon 0,03. Laman kanssa parhaan mallin selitysaste on 0,22 ja keskiarvon 0,12.

4.3.1.4 GARCH-mallin soveltuvuus kiinteistöindeksin ennustamiseen

Edellä esitetyn perusteella pystytään siis havaitsemaan, että mallilla on mahdollista saada melko hyviä ennusteita. Kuitenkin on huomattava, että ilman tietoa oikeista arvoista olisi parhaan mahdollisen ennusteen valitseminen paljon hankalampaa. Lisäksi tulosten pohjalta voi havaita selkeästi aikasarjan suuren vaikutuksen tuloksiin. Tässäkin on vaikea sanoa, mikä on sopiva valinta aikasarjoiksi. Ilman oikeiden havaintojen tuomaa turvaa, olisi esim. aikasarjan pituuden määrittäminen paljon vaikeampaa. Eikä se ole helppoa nytkään, sillä kuten tuloksista voi havaita, lamavuodet antavat paremman ennusteen kun ennusteajan oikeita arvoja ei käytetä ennusteen tekemisessä ja huonomman kuin niitä käytetään. Meillä ei valitettavasti ole

tarpeeksi pitkää aikasarjaa, jotta voisimme sanoa onko ero vain havaintojen lukumäärästä johtuvaa vai tuovatko lamavuodet jotain systemaattisesti erilaista malliin.

5 Yhteenveto

Kaikki kolme valittua tutkimussuuntaa osoittautuivat mahdollisesti sovellettaviksi. Jokaisella lähestymistavalla on kuitenkin omat ongelmansa, liittyen mallien kaksinaiseen luonteeseen: nyt haluttiin löytää sekä kiinteistöjen arvoja selittävät tekijät että pyrkiä luomaan niitä ennustava malli. Erittäin oleelliseksi asiaksi nousee myös käytettyjen aikasarjojen luonne – tulokset voivat olla hyvinkin erilaisia eri aikasarjoilla. Tämän projektin puitteissa emme päässeet tutkimaan kovinkaan montaa eri aikasarjaa, joten kovin pitkälle menevien johtopäätösten tekeminen on hankalaa.

Neuroverkolla saadaan luotua kohtuullisen hyviä malleja historiallisille arvoille. Ongelmana kuitenkin on, ettei malli tarjoa minkäänlaista taloudellista tulkintaa muuttujien todellisille riippuvuussuhteille. Tietysti herkkyysanalyysiä voidaan tehdä. Lisäksi neuroverkkojen käyttö kiinteistömarkkinoiden suunnan ennustamiseen voi olla haastavaa, eikä malli tarjoa kovin arvokasta tietoa markkinoiden kehitykseen vaikuttavista tekijöistä.

ARIMAX mallia käytettäessä löydettiin hyvin riippuvuussuhteet selitettävän aikasarjan ja selittäjien välillä. Pelkän AR-mallin käyttöön aikasarja vaikutti olevan aivan liian satunnaista, mutta ulkoisten selittäjien lisääminen paransi selitysasetta oleellisesti. Syvällisempi analyysi selittäjien merkitsevyydestä jäi kuitenkin tekemättä. Parhaimmillaan ARX-mallilla päästiin n. 50% selityssasteisiin.

GARCH toimii kohtuullisen hyvin ennustettaessa muutosten paikkaa, mutta mallin ongelmaksi tulee satunnaisuus. Satunnaisuuden takia mallin käytölle asettuu huomattavia rajoituksia – melkein minkäläinen tulos vaan voidaan saada aikaiseksi. Etenkin ennustetta tehtäessä pitää mieluummin tyytyä poikkeamien ajankohtien ennustamiseen kuin itse poikkeamien suuruuden käsittelyyn.

6 Pohdintoja

Projektin onnistuminen ja asetettuihin tavoitteisiin pääseminen ei tapahtunut aivan mutkattomasti ja tavoitteet täsmentyivätkin useasti projektin edetessä. Projekti eteni aluksi hyvin hallitusti, kun tehtäväkuvaa alettiin selvittää ja kirjallisuuskatsauksen tekeminen aloitettiin. Myöhemmin kuitenkin eri mallien sovellusvaiheessa päätettiin hajauttaa projekti kolmeen osaan. Tämä vaikutti oleellisesti projektin eri ryhmien väliseen kommunikaatioon, vaikka asia havaittiin jo riskien arviointi vaiheessa ja siihen yritettiin kiinnittää selkeästi huomiota. Projektin päämäärä pysyi selkeänä kuitenkin kaikilla osaryhmillä mielessä, vaikka kukin etenikin sitä kohti hiukan omaa polkuaan. Uskomme, että projektin lopputuloksena saimme tuotettua kohtalaisen hyvin uutta tietoa projektin asettajallemme juuri tutkimamme kolmen eri osa-alueen käytöstä tämän kaltaisen analyysin tekemisessä. Projektin aikataulu saneli kuitenkin melko ehdottomasti milloin työn teko tuli päättää. Projektin eteneminen tästä eteenpäin ja seuraavat vaiheet olisivat kuitenkin olleet kaikkiaan melko selvät. Mallien testaaminen erilaisella datalla olisi ollut seuraava etappimme.

Kaikkiaan ryhmämme piti projektia erittäin opettavaisena kokemuksena ja hyvänä näköalapaikkana siihen millaista projektien tekeminen oikeasti liikemaailmassa olisi. Voimmekin sanoa kurssin antaneen opetuksellisesti vähintään yhtä paljon projektityöskentelytaitoja kuin matemaattisten mallinnusten tekemiseen liittyviä taitoja.

Viitteet

- [Hay99] Simon Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall International Inc., 1999.
- [Kar98] Karolyi, Sanders: The Variation of Economic Risk Premiums in Real Estate Returns, Journal of Real Estate Finance and Economics 17:3 1998.
- [Kar05] Karakozova, Olga: Modelling and Forecasting Property Rents and Returns, 2005.
- [Bon03] Bond, Karolyi, Sanders: International Real Estate Returns: A Multifactor, Multicountry Approach, Dice Center Working Paper No. 2003-8.
- [Chi05] Chin, Fan: Autoregressive analysis of Singapore's private residential prices, Property Management, Volume 23, Number 4, 2005.
- [Sha00] Shachmurove, Y. ja D. Witkowska, Utilizing Artificial Neural Network Model to Predict Stock Markets, CARESS Working Paper #00-11, September, 2000.
- [Con98] Connellan, O. ja H. James, Estimated Realisation Price (ERP) by Neural Networks: Forecasting Commercial Property Values, Journal of Property Valuation and Investment, 1998, 16:1, 71-79.
- [Yu99] Yu, S., Forecasting and Arbitrage of the Nikkei Stock Index Futures: An application of Backpropagation Networks, Asian-Pacific Financial Markets, 1999, 6:4, 341-54.
- [Pin98] R.S. Pindyck & D.L. Rubinfeld: Econometric Models and Economic Forecasts, Fourth Edition, McGraw & Hill, New York, 1998.
- [Mat] Matlab help, <http://www.mathworks.com/support/>, 21.4.2007

Liitteet

Liite1: Neuroverkon matlab-koodi.

```
% Aikasarjan ennustaminen neuraaliverkon avulla
%
% parametrit
% x - piiloneuronien lukumäärät
% k - selitettävät muuttujat
% p - kuinka monta havaintoa otetaan selitettävistä muuttujista
%
% Esimerkiksi komennolla
%
% aikasarja([4 5], [1 2], [3 6])
%
% luodaan neuroverkko, jossa on 2 piilokerrosta, ensimmäisessä tasossa on 4
% neuronit ja toisessa tasossa 5 neuronit ja ulostulossa on 1 neuronit.
% Selitettäviä muuttujia on 1 ja 2, joista ensimmäisestä otetaan 3 edellistä havaintoa
% ja jälkimmäisestä 6 edellistä havaintoa. Neuroverkon luonnin jälkeen
% verkkoa opetetaan annetulla datalla kunnes etukäteen määritellyt ehdot
% toteutuvat, kts verkon parametrit. Opetuksessa käytetty data ei sisällä kaikkia havaintoja, kts
% muuttuja M.

function v = aikasarja(x, k, p)

% Ladataan data tiedostosta

ts = load('data2.txt');
ts = ts';

% osalle datalle tehdään muunnoksia

ts(1,1:end) = log(ts(1,1:end));
ts(2,1:end) = log(ts(2,1:end));
ts(4,1:end) = log(ts(4,1:end));

% L kertoo kuinka monta syötettä
L = 0;
for i = 1:size(p,2)
    L = L + p(i);
end

% K kertoo kuinka monta eri muuttujaa
K = size(k,2);
% M opetusdatan osuus
M = 180;
% koko datan pituus
E = size(ts,2);

% luodaan taulukot, joita käytetään opetuksessa ja
% ennustamisessa. Lopuksi luodaan taulukot koko datasta
```

```

opetus = [];
%% %%
for i = 1:K,
    for j = 1:p(i),
        opetus = [opetus; ts(k(i), j:(M+j-1))];
    end
end
kohde = [ts(1,1+p(1):(M+p(1)))];

testi = [];
for i = 1:K,
    for j = 1:p(i),
        testi = [testi; ts(k(i), (M+j-1):(E-p(1)+j-1))];
    end
end
oikea = [ts(1,(M+p(1)):E)];

koko = [];
for i = 1:K,
    for j = 1:p(i)
        koko = [koko; ts(k(i), j:(E-p(1)+j-1))];
    end
end
tavoite = [ts(1,(p(1)+1:E))];

% Aktivointi funktiot

F1 = {'tansig', 'purelin'};
F2 = {'tansig', 'purelin', 'purelin'};

if(length(x) > 1)
    F = F2;
else
    F = F1;
end

M = [min(opetus') max(opetus')];
N = [x 1];

% luodaan neuroverkko
net = newff(M, N, F, 'trainsg', 'learnsgdm', 'mse');
net;

% Alustetaan input- ja output-tason painot sekä biasit
if (length(x) > 1)
    net.IW{1} = 0.001*randn([x(1) L]);
    net.LW{2} = 0.001*randn([x(2) x(1)]);
    net.LW{3,2} = 0.001*randn([1 x(2)]);
    net.b{1} = 0.001*randn([x(1) 1]);
    net.b{2} = 0.001*randn([x(2) 1]);
    net.b{3} = 0.001*randn;
else
    net.IW{1} = 0.001*randn([x(1) L]);
    net.LW{2} = 0.001*randn([1 x(1)]);
    net.b{1} = 0.001*randn([x(1) 1]);
    net.b{2} = 0.001*randn;
end

```

```

end

% Alustetaan verkon parametrit
net.trainParam.show = 50;
net.trainParam.lr = 0.01;
net.trainParam.epochs = 100000;
net.trainParam.goal = 1e-6;
net.trainParam.mc = 0.9;
net.trainParam.min_grad = 1e-28;
net.trainParam.mu_max = 1e300000;
net.trainParam.minstep = 1.0e-132;

% Opetus
[net,tr,Y,E0]=train(net,opetus,kohde);

% Ennustetaan verkolla
[ennuste1,Pf1,Af1,E1,perf1] = sim(net, opetus);
[ennuste2,Pf2,Af2,E2,perf2] = sim(net, testi);
[ennuste3,Pf3,Af3,E3,perf3]= sim(net, koko);

%piirretään kuvaajat
figure(2);

subplot(2,2,1);
plot(kohde);
%h = legend('off');
title('opetusjoukon ennuste ja oikea arvo');
hold on
plot(ennuste1,'r+');
hold off

subplot(2,2,2);
plot(oikea);
title('testijoukon ennuste ja oikea arvo');
hold on
plot(ennuste2,'r+');
hold off

subplot(2,2,3);
plot(kohde - ennuste1);
title('opetusjoukon ennusteen ja oikean arvon erotus');

subplot(2,2,4);
plot(oikea - ennuste2);
title('testijoukon ennusteen ja oikean arvon erotus');

figure(3)
plot(oikea);
title('testijoukon ennuste ja oikea arvo');
hold on
plot(ennuste2,'r+');
hold off

% Virheet, jotka lasketaan käyttämällä verkkoa
ennuste = sim(net, opetus);
opetus_svirhe = var(kohde - ennuste) / var(kohde)

ennuste = sim(net, testi);

```

```
testi_svirhe = var(oikea - ennuste) / var(oikea)
```

```
ennuste = sim(net, koko);
```

```
koko_svirhe = var(tavoite - ennuste) / var(tavoite)
```

```
figure(4);  
plot(exp(tavoite));  
title('Koko joukon ennuste ja oikea arvo');  
ylabel('EPRA indeksin arvo');  
xlabel('Havainnot');  
hold on;  
plot(exp(ennuste3), 'r--');  
legend('Oikea arvo', 'ennuste');  
hold off;  
axis tight  
print -f4 -dtiff 'ajo1a'
```

```
figure(5);  
plot(exp(tavoite)-exp(ennuste3));  
title('Tavoitearvon ja ennustetun arvon erotus');  
xlabel('Havainnot');  
ylabel('Absoluuttinen virhe');
```

```
figure(6);  
plot((exp(tavoite)-exp(ennuste3))./exp(tavoite)*100);  
title('Virhe prosentteina');  
xlabel('Havainnot');  
ylabel('Prosentuaalinen virhe');  
axis tight  
print -f6 -dtiff 'ajo1b'
```

```
% tulostetaan näytölle painokertoimet
```

```
net.IW{1};
```

```
net.LW{2};
```

```
net.b{1};
```

```
net.b{2};
```

```
% lasketaan selityssaste
```

```
keskiarvo_y = mean(ts(1,1:end))
```

```
ss_yy = 0;
```

```
for i = 1:size(ts,2),
```

```
    ss_yy = ss_yy + (ts(1,i) - keskiarvo_y)^2;
```

```
end
```

```
sse = 0;
```

```
for i = 1:length(tavoite),
```

```
    sse = sse + (tavoite(i) - ennuste3(i))^2;
```

```
end
```

Liite 2: GARCH-malli

Mallin selittävät muuttujat: osakeindeksi, 3kk korko
Aikasarjan data vuosilta: 31.1.1995–28.2.2007

Mean: ARMAX(1,0,2); Variance: GARCH(1,1)

Conditional Probability Distribution: T
Number of Model Parameters Estimated: 8

Parameter	Value	Standard Error	T Statistic
C	0.0079897	0.0031147	2.5652
AR(1)	-0.019918	0.079301	-0.2512
Regress(1)	0.34156	0.068067	5.0180
Regress(2)	0.067837	0.048913	1.3869
K	7.354e-005	0.00018776	0.3917
GARCH(1)	0.91409	0.1655	5.5231
ARCH(1)	0.044494	0.07033	0.6327
DoF	5.5961	2.7755	2.0163

Parhaan yksittäisen ennusteen selityssaste =

0.2025

Ennusteiden keskiarvon korrelaatio oikeiden indeksin arvojen kanssa =

0.5599

Ennusteiden keskiarvon selityssaste =

0.3136

Mallin selittävät muuttujat: osakeindeksi, 3kk korko
Aikasarjan data vuosilta: 31.1.1995–28.2.2005

Mean: ARMAX(1,0,2); Variance: GARCH(1,1)

Conditional Probability Distribution: T
Number of Model Parameters Estimated: 8

Parameter	Value	Standard Error	T Statistic
C	0.0095135	0.0030599	3.1091
AR(1)	-0.013744	0.08196	-0.1677
Regress(1)	0.084393	0.064459	1.3093
Regress(2)	0.043977	0.043199	1.0180
K	0.00014689	0.0003824	0.3841
GARCH(1)	0.97274	0.069877	13.9206
ARCH(1)	0.027264	0.17742	0.1537
DoF	2.3404	0.54583	4.2878

Parhaan yksittäisen ennusteen selityssaste =

0.1587

Ennusteiden keskiarvon korrelaatio oikeiden indeksin arvojen kanssa =

-0.1712

Ennusteiden keskiarvon selityssaste =

0.0293

Mallin selittävät muuttujat: osakeindeksi, 3kk korko
Aikasarjan data vuosilta: 31.1.1990-28.2.2007

Mean: ARMAX(1,0,2); Variance: GARCH(1,1)

Conditional Probability Distribution: T
Number of Model Parameters Estimated: 8

Parameter	Value	Standard Error	T Statistic
C	0.004549	0.0026201	1.7362
AR(1)	0.058985	0.070261	0.8395
Regress(1)	0.00033451	0.040992	0.0082
Regress(2)	0.41948	0.059296	7.0744
K	0.00051294	0.0019504	0.2630
GARCH(1)	0.60267	1.4671	0.4108
ARCH(1)	0.027222	0.088158	0.3088
DoF	10.864	7.5897	1.4315

Parhaan yksittäisen ennusteen selityssaste =

0.2336

Ennusteiden keskiarvon korrelaatio oikeiden indeksin arvojen kanssa =

0.2170

Ennusteiden keskiarvon selityssaste =

0.0471

Mallin selittävät muuttujat: osakeindeksi, 3kk korko
Aikasarjan data vuosilta: 31.1.1990-28.2.2005

Mean: ARMAX(1,0,2); Variance: GARCH(1,1)

Conditional Probability Distribution: T
Number of Model Parameters Estimated: 8

Parameter	Value	Standard Error	T Statistic
C	0.0073068	0.002961	2.4677
AR(1)	0.050778	0.078051	0.6506
Regress(1)	0.031717	0.040726	0.7788
Regress(2)	0.047014	0.072066	0.6524
K	0.0011346	0.001878	0.6042
GARCH(1)	0.27207	1.1524	0.2361
ARCH(1)	0.067363	0.1101	0.6118
DoF	6.6538	3.5791	1.8591

Parhaan yksittäisen ennusteen selityssaste =

0.2188

Ennusteiden keskiarvon korrelaatio oikeiden indeksin arvojen kanssa =

0.3457

Ennusteiden keskiarvon selityssaste =

0.1195