VATT-TUTKIMUKSIA
67
VATT-RESEARCH REPORTS

Lauri Kyllönen - Tarmo Räty

ASUNTOJEN HINTA-LAATUSUHDE JOENSUUSSA, SEMIPARAMETRINEN ESTIMOINTI

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2000

ISBN 951-561-330-2

ISSN 0788-5008

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Hämeentie 3, 00530 Helsinki, Finland

Email: tarmo.raty@vatt.fi
lauri.kyllonen@joensuu.fi

J-Paino Oy

Helsinki, 2000

KYLLÖNEN, LAURI ja RÄTY, TARMO: ASUNTOJEN HINTA-LAATUSUHDE JOENSUUSSA, SEMIPARAMETRINEN ESTIMOINTI. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research. 2000, (B, ISSN 0788-5008; No 67), ISBN 951-561-330-2.

Tiivistelmä: Asuntomarkkinat käsitetään usein alueen tai asuntotyyppien mukaan luokiteltuna kokonaisuutena, jonka päällimmäisenä kysymyksenä on hintakehitys. Kuitenkin asunnon myynti tai hankintapäätökseen liittyy muitakin määritteitä, kuten asunnon pinta-ala, varustelutaso, yleiskunto, ikä tai liikenneyhteydet. Yksikään näistä tekijöistä ei pysty yksin selittämään markkinoilla määräytyvää hintaa vaan ostajat pyrkivät löytämään markkinoilta näiltä ominaisuuksiltaan parhaan mahdollisen yhdistelmän. Olemme tutkineet tässä paperissa Joensuun asuntomarkkinoiden hintoja ja myytyjen vapaarahoitteisten kerros- ja rivitaloasuntojen ominaisuuksia. Käytetty kauppakohtainen aineisto kattaa noin 40 % kaikista vuosina 1985-1997 myydyistä asunnoista.

Käyttämämme hedoninen hintafunktio selittää asuntojen hintoja niiden ominaisuuksilla. Funktion olemassaolo voidaan tietyin edellytyksin osoittaa, mutta sen ominaisuuksista on olemassa hyvin vähän yleispätevää tietoa. Pääsääntöisesti kaikissa tutkimuksissa funktiolle on annettu jokin parametrinen muoto. Tässä tutkimuksessa pyrimme estimoimaan funktion ilman merkittäviä ex ante oletuksia, joten olemme yhdistäneet parametriseen hedoniseen malliin ei-parametrisen splinifunktion, ranskalaisen käyrän. Tätä malliyhdistelmää kutsutaan semiparametriseksi funktioksi. Koska estimoitu funktio ja sen estimointimenetelmä ovat taloustieteellisessä keskustelussa vakiintumattomia, olemme sisällyttäneet tekstiin suhteellisen pitkän selostuksen splineistä ja sen "roughness penalty" estimaattorista.

Estimoinnit Joensuun asuntomarkkinoilla viittaavat asunnon iän suureen merkitykseen uusilla asunnoilla. Uuden asunnon ostaja maksaa ensimmäisen 15 vuoden aikana yli 100 000 markan preemion pelkästään asunnon iästä. Samalla asuntomarkkinat näyttävät jakautuneen kahteen osaan asunnon iän suhteen, sillä ennen 80-lukua rakennetuilla asunnoilla iän merkitys hintaan on olennaisesti pienempi. Asuntomarkkinoilta voidaan myös paikallistaa noin 4 % kaupankäynnistä kattava alamarkkina, jonka piirissä asunnot ovat pienempiä, huonokuntoisempia, hoitovastikkeeltaan kalliimpia, kauempana keskustasta ja rannasta sekä suhteellisesti kalliimpia kuin muut asunnot. Jos ostajalla ei ole varaa muuhun kuin näihin absoluuttisesti halpoihin mutta laatuunsa nähden kalliisiin asuntoihin, joutuu hän paradoksaaliseen "köyhällä ei olisi varaa halpaan" tilanteeseen. Alueellisesti nämä markkinat kattavat pahimmillaan yli 15 % kaupankäynnistä.

Ei-parametristen menetelmien valinnalle esitämme kolme perusteltua syytä. Ensinäkin ne ovat tehokas ja selkeä tapa visualisoida mitä tahansa aineistoa. Toiseksi splini on tehokas väline hyvin epälineaaristen relaatioiden etukäteistarkastelussa. Lopullinen malli voidaan perustaa johonkin yksinkertaistukseen. Kolmanneksi ei kuitenkaan vähiten merkityksellisenä, splinillä on mahdollista estimoida muutossuuntia ja -nopeuksia aineistosta, jossa mielenkiinnon kohteena olevan tekijän vaikutus selitettävään tekijään on epälineaarinen tai hyvin poukkoileva.

Asiasanat: Asuntomarkkinat, Alamarkkinat, Ranskalainen käyrä, Splini-funktio, Ei-parametrinen estimointi, Semi-parametrinen estimointi, Joensuu

Abstract: Whereas most of the hedonic price function studies use specific parametric forms to describe the impact of the properties of the goods on their prices, the nonparametric approach used in this paper allows the variable freely alternate the direction and the curvature of the impact at any observed level. The estimated relation between the property of the good and its price is called a French curve, or more formally a spline function. Augmented with the traditional parametric model, we estimate a semiparametric hedonic function.

We have analyzed the privately financed housing markets of the city of Joensuu using the French curve approach. Our results show, that the housing markets in Joensuu work in general rather efficiently, so that the price of a dwelling and its quality usually matches on the marketplace. However, a more detailed analysis on the markets revealed submarkets, where dwellings are exorbitant. The nominal prices of these dwellings are usually low and the submarkets cover 15 percent of certain suburbs.

It was found that the nonparametric variable, the age of dwelling, has a strong impact on prices of the dwellings constructed durin 1980's and 1990's. The older dwelling base experienced more moderate price behavior. This result was insensitive to the period of estimation. Thus, the exogenous shocks having impact on prices of the newly built dwellings do not necessarily spread over the whole markets, but they remain as a premium for the buyer.

We have found three justifications for the use of nonparametric methods. First, they are an efficient and easy way to visualize any data set. Second, the spline function is an efficient tool to preanalyze inherently nonlinear relations. The final model may be based on the simplified model. Third but not least, the spline function allows us to estimate the slopes and their changes from highly nonlinear or bumpy data.

Keywords: Housing markets, Submarkets, French curve, Spline function, Nonparametric estimation, Joensuu

Sisältö

Yhteenveto

Asuntomarkkinoiden tutkimuksessa joudutaan usein ottamaan huomioon tekijöitä, joilla on pienempi merkitys muussa markkinatutkimuksessa. Ensinnäkin harvalla hyödykkeellä on niin suuri paino kuluttajan budjetissa kuin asunnolla. Ostaja joutuu punnitsemaan erilaisia asumisratkaisuja suhteessa jäljelle jäävään ostovoimaan. Yleensä pitkäaikaisena ratkaisuna asunto vaikuttaa edelleen niin työhön kuin vapaa-aikaankin liittyviin valintoihin. Toiseksi jo pelkästään asuntoon liittyvät tekijät muodostavat oman valintaongelmansa. Kun kulutushyödykkeet ovat usein ominaisuuksiltaan yhtenäisiä, kahta täysin samanlaista asuntoa voi olla vaikea löytää. Kuluttajan täytyy ensin löytää mekanismi, jossa vertaillaan hyvin heterogeenista asuntotarjontaa, ja sen jälkeen suhteuttaa valintamahdollisuudet (hyvin pitkällä ajanjaksolla) muihin mahdollisuuksiin. Tässä tutkimuksessa lähdemme liikkeelle siitä, että kuluttaja tarvitsee uuden asunnon ja on päättänyt siihen sijoittaa. Analysoimme asuntomarkkinoiden hinta-laatu suhteita Joensuussa mallilla, joka poistaa muiden kulutusvaihtoehtojen vaikutuksen, ja mahdollistaa painotuksen asuntojen laatutekijöiden tarkasteluun.

Mikäli asuntojen laadun mukaisesti lasketussa hinnassa ja markkinahinnassa havaitaan poikkeavuuksia, sitä voidaan selittää paikallisilla tai jonkin muun asunnon ominaisuuden suhteen keskittyneillä asuntojen alamarkkinoilla. Alamarkkinoiden havainnointi edellyttää asuntojen ominaisuuksien muutoksille hyvin herkkää mittaria. Mikäli asuntojen markkinahinnan riippuvuus sen ominaisuuksista estimoidaan parametrisena, malli ei pysty sopeutumaan voimakkaisiin ja ennalta tuntemattomiin epälineaarisiin muutoksiin ominaisuuksien vaihtelussa. Sen vuoksi tässä tutkimuksessa estimoidaan hedoninen funktio osin ei-parametrisena. Siinä vaikutusta ei sidota etukäteen mihinkään funktiomuotoon. Esimerkiksi asunnon iän vaikutusta sen hintaan ei oleteta tasaisesti laskevaksi tai paraabeliksi, vaan haemme havaintojen perusteella sille sopivan muodon. Lähestymistapa sallii vaikkapa korjausrakentamisen tai eri aikakausina rakennettujen asuntojen vaikutuksen erittelyn ilman vahvoja ennakko-oletuksia ilmiön luonteesta.

Käyttämäämme ei-parametrista splini-tekniikkaa kutsutaan myös nimellä ranskalainen käyrä, sillä matemaattisesti pitkähkön estimointiproseduurin tuloksena saadaan havaintoja pehmeästi mukaileva käyrä—havaintojen kuvaus ranskalaisittain. Käyrä voi vapaasti muuttaa kulkusuuntaansa valituissa selittävän muuttujan solmukohdissa, sen mukaan kuinka solmun ja sen ympäristön havainnot antavat paremman sovitteen. Periaatteessa hyvinkin epälineaarisia relaatioita voitaisiin kuvata splinillä, mutta menetelmä vaatii suhteellisen suuren määrän havaintoja stabiloituakseen. Käytännössä havaintoja pitää olla useita satoja ja solmukohtien lukumäärä ei saisi ylittää muutamaa kymmentä. Molemmat tekijät vaikuttavat myös laskennalliseen vaativuuteen. Olemme tässä työssä pystyneet estimoimaan pöytäkoneella semiparametrisen mallin, joka yhdistää niin parametrisia kuin ei-parametrisiakin muuttujia samaan malliin yli 1700 havainnolla ja 46 solmukohdalla. Numeerisiin ongelmiin orientoituneilla ohjelmistoilla voidaan käsitellä huomattavasti laajempiakin ongelmia, joten mallien väitetty laskennallinen raskaus on nykytekniikalla vahvasti väistymässä.

Splini-funktio ei ole tarkkaan ottaen ei-parametrinen funktio, vaan muuttuvaparametrinen polynomi. Ranskalainen käyrä saadaan aikaiseksi valitsemalla jokaiselle solmuvälille kuutiopolynomin parametrit siten, että vierekkäisten solmuvälien polynomien arvo sekä kaksi ensimmäistä derivaattaa ovat solmussa yhtäsuuria. Käyttämämme malli on lisäksi määritelty siten, että "yhteinen" kuutiopolynomi yli koko havaintovälin saadaan mallin erikoistapauksena. Tämä mahdollistaa tilastollisen testauksen parametrisen mallin ja splinin välillä.

Aineisto koostuu Suomen kiinteistövälittäjien liiton jäsenyritysten kautta Joensuun kaupungin alueella vuosina 1985 - 97 käydyistä kerros- ja rivitaloasuntokaupoista. Yhteensä havaintoja on 2023 joista 1713 riittävän täydellisiä tutkimukseen. Selitettävä muuttuja on asunnon velallinen kauppahinta. Selittävinä muuttujina käytetään asunnon pinta-alaa, sijaintikerrosta, yhtiövastiketta (sisältäen sekä hoito, että rahoituserät), keskustaetäisyyttä, tien ja rannan läheisyyttä sekä ei-parametrisina muuttujina aikatrendiä ja asunnon ikää. Valitsemalla aikatrendin ei-parametriseksi pystymme estimoimaan tarkasti havaintoperiodista johtuvan hintavaihtelun. Asunnon iällä haluamme tutkia miten asunnon ikään liittyvät ominaisuudet näkyvät hintakehityksessä.

Estimointituloksissa havaitsimme, että asunnon iällä sen myyntihinnan selittäjänä on hyvin vaihteleva merkitys. Rakennuskanta Joensuussa näyttää jakaantuneen kahteen osaan. Uusiin 80- ja 90-luvuilla rakennettuihin asuntoihin iällä on hyvin voimakas merkitys. Ensimmisenä 15 vuotenaan nämä asunnot menettävät hinnastaan karkeasti 100 000 markkaa pelkästään ikätekijän vaikutuksesta. Tähän tulee lisätä vielä mahdolliset rahoitusvastikkeet sekä suoraan ikään liittyvä arvon aleneminen, kuten asunnon kunnon heikkeneminen. Toisaalta rakennuskannan vanhalla osalla, ennen 80-lukua rakennetuilla asunnoilla, ikä ei ole enää ratkaiseva tekijä asunnon hinnassa. Asunnot mielletään jo vanhoiksi ja hinta määräytyy paremminkin muiden asunnon ominaisuuksien perusteella. Tälläisella markkinatilanteella uustuotantoon kohdistuvat hintapaineet jäävät helposti uusasunnon ostajan kannettavaksi, sillä toimivat markkinat painavat asunnon hinnan ajan myötä laatutasoa vastaavaksi.

Asuntojen hinta-laatusuhteita tarkasteltiin alamarkkinahypoteesin kannalta. Mikäli markkinoiden hinnoittelu jollakin estimoidulla asuntojen laatutasolla on selvästi yli tai alle laatutason edellyttämän, asuntokauppa tällä laatutasolla muodostaa oman alamarkkinansa. Joensuusta estimoimallamme mallilla havaitsimme, että hyvin vaatimattomista asunnoista maksettiin markkinoilla ylihintaa. Tämän alamarkkinan merkitys on kokonaisuutta katsoen pieni, mutta Rantakylän ja Mutalan alueella 10 - 20 % asuntokaupasta tapahtuu tällä segmentillä. Havainnon tekee poliittisesti ongelmalliseksi se, että markkinoiden kiristyminen kohdistuu halpoja pienasuntoja tarvitseviin joko suoraan asunnon ylihintana ominaisuuksiin nähden tai korkeampana vuokratasona. Ilmeisesti sosiaalinen asuntotuotanto Joensuussa ei pysty näitä markkinoita riittävästi tukemaan.

1 Johdanto

[1]Asuntomarkkinoiden tutkimuksessa joudutaan usein ottamaan huomioon tekijöitä, joilla on pienempi merkitys muussa markkinatutkimuksessa. Ensinnäkin harvalla hyödykkeellä on niin suuri paino kuluttajan budjetissa kuin asunnolla. Ostaja joutuu punnitsemaan erilaisia asumisratkaisuja suhteessa jäljelle jäävään ostovoimaan. Yleensä pitkäaikaisena ratkaisuna asunto vaikuttaa edelleen niin työhön kuin vapaa-aikaankin liittyviin valintoihin. Toisaalta jo pelkästään asuntoon liittyvät tekijät muodostavat oman valintaongelmansa. Kun kulutushyödykkeet hyödykkeet ovat usein ominaisuuksiltaan yhtenäisiä, kahta täysin samanlaista asuntoa voi olla vaikea löytää. Kuluttajan täytyy ensin löytää mekanismi, jossa vertaillaan hyvin heterogeenista asuntotarjontaa, ja sen jälkeen suhteuttaa valintamahdollisuudet hyvin pitkällä ajanjaksolla muihin mahdollisuuksiin. Tässä tutkimuksessa lähdemme liikkeelle siitä, että kuluttaja tarvitsee uuden asunnon ja on päättänyt siihen sijoittaa. Analysoimme asuntomarkkinoiden hinta-laatu suhteita Joensuussa välineillä, jotka ottavat huomioon sekä valinnan asunnon ja muun kulutuksen välillä että asuntojen laatutekijät.

Tarkastelemme kuluttajan resurssien allokointia asumisen ja muun kulutuksen välillä täydellisen kilpailun tasapainoratkaisuna. Tämä mahdollistaa keskittymisen mielestämme olennaisempaan ongelmaan eli siihen, miten asuntomarkkinoiden heterogeenisuus vaikuttaa niiden tehokkuuteen. Hyvin homogeenisten hyödykkeiden markkinat toimivat usein tehokkaasti, sillä kuluttajat voivat helposti etsiä korvaavan tuotteen ja tarjoajat myydä tuotteensa toiselle. Asuntomarkkinoilla tilanne on vaikeampi. Markkinat ovat jatkuvat ja yhtenäiset vain, jos asuntojen ominaisuuksien välinen keskinäinen korvattavuus on helppoa ja eri ominaisuusyhdistelmät voidaan asettaa laadun suhteen samalle asteikolle. Täydellinen informaatio kaupankäynnin kohteena olevista asunnoista on perusedellytys jälkimmäiselle. Vaikka tämä voikin olla jossain määrin realistinenkin oletus, niin hyvin kauaskantoiseen ostopäätökseen voi sisältyä vaikeasti korvattavissa olevia toiveita asunnosta. Tällaisessa tapauksessa ostajan kannalta markkinat pienenevät ja asunnon hinta-laatusuhde voi poiketa merkittävästi yleisestä markkinatilanteesta.

Ostajalle asunnon laatu voi olla voimakkaan sitoutunut johonkin sen ominaisuuteen. Tyyppiesimerkki lienee asunnon muiden ominaisuuksien pienentynyt merkitys, kunhan se vain sijaitsee arvostetulla asuntoalueella. Ostaja ilmaisee laatuarvionsa viimekädessä maksamallaan hinnalla. Alueen ja asunnon hinnan välinen riippuvuus tunnetaan yleisesti, mutta voivatko muut asuntojen ominaisuudet synnyttää sa-

[1]Olemme kiitollisia tutkijakollegoillemme ja erityisesti Juha Alholle, Seppo Laaksolle ja Heikki Loikkaselle rakentavista kommenteista työn edetessä. Tämä ei kuitenkaan vapauta meitä tekstiin jääneistä virheistä.

mantapaista keskittymistä yhteen laatutasoon? Tämä on yksi tärkeä kysymys, johon pyrimme tässä tutkimuksessa vastaamaan etsimällä asuntomarkkinoiden alamarkkinoita. Alamarkkina[2] syntyy kun ostajat ja myyjät ovat halukkaita käymään kauppaa vain jollain suppealla laatutason vaihtelulla siten, että hinta-laatu suhde poikkeaa merkittävästi ympäristöstään. Tämä tekee alamarkkinat tärkeäksi myös yhteiskunnallisen päätöksenteon kannalta. Suhteellisesti kalliisiin alamarkkinoihin voidaan vaikuttaa suuntaamalla asuntopolitiikkaa, joko paikallisesti tai valtakunnallisesti, tasaamaan hintapiikkejä. Erityisesti aluetasolla sosiaalinen asuntotuotanto voi näin tukea markkinakapeikkojen aukaisemista.

Asuntopolitiikan vaikutusten arviointi edellyttää alamarkkinoiden tuntemusta, sillä heikko korvattavuus näkyy Rothenberg *et al.* [RGBP91] mukaan myös laatuun liittymättömien (eksogeenisten) muuttujien vaikutuksessa. Esimerkiksi korkovähennysjärjestelmän poistaminen vaikuttaisi vähemmän hyvin vaikeasti korvattavilla alamarkkinoilla kuin joustavilla markkinoilla, jossa asunnon suhteellinen kallistuminen siirtäisi ostajia joko alemmalle laatutasolle tai hinnat laskisivat. Jo toteutetun esimerkin eksogeenista vaikutuksesta asuntomarkkinoille tarjoaa hallituskaudella 1995 - 99 tehdyt leikkaukset yksityisteiden ylläpidon avustuksiin. Muutoksesta kärsivät ne haja-asutusalueiden asukkaat, joille vaikkapa luonnon rauha on vaikeasti korvattava ominaisuus. Haja-asutusalueiden asuntojen markkinahinta pyrkivät tässä tilanteessa laskemaan. Etukäteen ajateltuna ainakin asunnon sijainti ja ikä ovat vaikeasti korvattavia ominaisuuksia, jotka voivat synnyttää alamarkkinoita.

Markkinoiden tehokkuuden heikkeneminen ja tarve tukea niiden toimintaa voidaan myös kytkeä transaktiokustannusteoriaan [Wil81]. Sen mukaan ihmisiä ohjaa perinteisiin uusklassisiin malleihin verrattuna heikompi ongelman ratkaisukyky eli rajoitettu rationaalisuus ja ainakin jotkut ovat taipuvaisia opportunistiseen käyttäytymiseen eli lähinnä petkuttamiseen. Markkinoiden tehokkuus riippuu transaktion piirteistä. Niistä tärkeimmät ovat epävarmuus, transaktioiden tiheys ja niiden kohteen erityispiirteet. Kaikki nämä piirteet ovat hyvin edustettuina asuntomarkkinoilla: tuotteen laatu ei ole aina varma, jotkut ostajat voivat olla epävarmoja maksukyvystään, yksittäinen asunto on harvoin kaupankäynnin kohteena ja asunnoilla voi olla lukuisia ominaispiirteitä.

Olemme estimoineet asuntomarkkinoiden hinta-laatu suhdetta Joensuussa. Käytössä ollut aineisto kattaa noin 2000 Suomen kiinteistövälittäjäin liiton jäsenyritysten kautta kaupattua kerros- ja rivitaloasuntoa vuosina 1985 - 1997. Käyttämämme hintamalli on periaatteessa hyvin yksinkertainen. Siinä selitetään asunnosta mak-

[2]Tässä tutkimuksessa käytettyä alamarkkina-käsitettä on esitelty alunperin Rothenberg *et al.* [RGBP91]. Olemme joutuneet sovittamaan määrittelyä paremmin suomalaiseen asuntomarkkinakäytäntöön sopivaksi.

settua hintaa sellaisilla ominaisuuksilla kuin asunnon pinta-ala, vastike, kerros ja sijainti suhteessa kaupungin keskustaan, rantaan ja häiritsevään liikenteeseen. Eri alamarkkinoiden rajat havainnoidaan vertaamalla keskenään asunnoista markkinoilla maksettua hintaa ja käyttämämme mallin mukaisesti laskettua asunnon ominaisuuksia kuvaavaa laatuhintaa. Systemaattiset poikkeamat näissä hinta-laatu suhteissa kertovat markkinahäiriöistä, maksetuista yli- tai alihinnoista verrattuna aineiston keskimääräiseen tasoon. Koska mallilla laskettavan laatuarvion merkitys on hyvin keskeinen tulosten luotettavuuden kannalta, olemme halunneet olla hyvin tarkkoja keskeisten muuttujien spesifioinnissa. Tämä on edellyttänyt tavallisesta hieman poikkeavaa estimointimenetelmää, joka ei välttämättä ole tuttu edes ammattiekonomisteille. Toivomme, että estimaattorin yksityiskohtainen kuvaus ei hämää lukijaa, sillä varsinaiset tulokset ovat maallikollekin jopa helpommin tulkittavissa kuin perinteiset regressiomallin tulokset.

Tarkoituksenamme on ollut asuntomarkkinoiden alamarkkinoiden tutkimisen rinnalla esittää sellainen matemaattinen ja tilastollinen pohja estimointimenetelmällemme, jonka perusteella ekonometrikko pystyy operoimaan ja arvioimaan tuloksia. Niinpä estimointitekniikkaan keskittyvät kappaleet 2.2-2.7 vaativat lukijalta perehtyneisyyttä lineaarialgebraan ja laskenta-algoritmeihin. Allekirjoittaneiden arvostelukykyyn luottaen lukija voi siirtyä suoraan ehkä kappaleen 2.1 luettuaan, Joensuun asuntomarkkinoihin kappaleessa 3. Siinä on esitetty aineisto ja empiirisen työn tulokset: semiparametrinen hedoninen hintafunktio kahdelle eri aikaperiodille ja asuntojen alamarkkinat Joensuussa. Kappaleessa 4 vertailemme estimointimenetelmäämme perinteisempiin menetelmiin. Viimeisessä kappaleessa esitämme muutamia johtopäätöksiä asuntomarkkinoiden toiminnasta ja käytetyn estimointimenetelmän ominaisuuksista.

2 Ranskalaisen käyrän estimointi

2.1 Hedoninen hintateoria asuntomarkkinoilla

Uusklassisessa teoriassa asuntomarkkinoiden hinnanmuodostusta kuvataan kysynnän ja tarjonnan tasapainon toteuttavalla hintafunktiolla. Se luonnehtii niitä ominaisuus-hinta yhdistelmiä, joissa markkinoilla olevat asunnot vaihtavat omistajaansa siten, että täydellisen tietämyksen omaavien kaupan osapuolten tavanomaiset uusklassiset tavoitefunktiot maksimoituvat. Tasapainossa ostajan maksuhalukkuuden ilmaiseva tarjousfunktio ja myyjän hintapyyntöfunktio sivuavat. Heterogeenisilla markkinoilla sivuamispisteitä on lukuisia ja niistä muodostuva ura ilmaisee ominaisuuksien varjohinnat. Tätä kutsutaan hedoniseksi hintafunktioksi. Tasapainossa ostaja ei voi lisätä hyötyään, eikä myyjä voittoaan valitsemalla toisin. Lisäksi tarjolla olevien ominaisuuskombinaatioiden valintajoukko täytyy olettaa jatkuvaksi; ominaisuuksien on oltava vapaasti korvattavia, jotta rajasuureiden vertailuun perustuva maksimoiva käyttäytyminen olisi mahdollista.

Jatkuvuusoletus on herättänyt kritiikkiä, sillä käytännössä asunnon ominaisuudet valitaan diskreetistä joukosta ja lopullinen valinta kohdistuu vain yhteen asuntoon. Niinpä vaihtoehtoisena lähestymistapana on käytetty diskreetin valinnan malleja. Tasapainomalli ei myöskään pysty kuvaamaan toimijoiden uskomuksia parhaista saavutettavissa olevista osto- ja myyntitarjouksista. Kuitenkin hintapyynnöt ja ostotarjoukset riippuvat näistä uskomuksista [Pal91, sivut 116-119]. Epätäydellisillä markkinoilla hedoninen hintafunktio pitäisi korvata optimivalintoja kuvaavalla hinta-laatu tehokkuuskäyrällä, jolta rationaalisesti käyttäytyvä kuluttaja valitsee asuntonsa [Phl83, sivu 206].

Alamarkkinoiden esiintyminen merkitsee Rosenin [Ros74] hedonisen teorian oletusten kumoutumista. Käytämme tässä tutkimuksessa tätä ristiriitaa hyväksemme etsimällä poikkeamia hedonisen mallin osoittamasta käyttäytymisestä. Emme rakenna malleja alamarkkinoille, vaan rakennamme koko aineiston kattavan ekonometrisen kuvauksen hedonisella hintafunktiolla. Mallin huono selityskyky jollakin laatutasolla tai ominaisuusyhdistelmällä voidaan tulkita alamarkkinoiksi tai mallin spesifikaatiovirheeksi. Niinpä oikea markkinoiden laajuus, mahdolliset muuttujien multikollineaarisuudet tai proxy-muuttujien käyttö sekä funktiomuodon valinta ovat keskeisiä tekijöitä tuloksia arvioitaessa.

2.1.1 Hintafunktion muodon valinta

Rosenin teoriassa kauppahinta riippuu tiukasti ottaen vain tuotteen ominaisuuksien kysynnästä ja tarjonnasta. Teoria kertoo kuitenkin kovin vähän itse funktiosta ja sen ominaisuuksista. Yleensä tätä ongelmaa on ratkottu valitsemalla jokin hyvin joustava funktiomuoto estimoinnin pohjaksi ja annettu havaintojen määritellä mahdollisimman pitkälle kunkin ominaisuuden vaikutustapa. Tutkimuksissa on havaittu vain pieniä eroja eri funktiomuotojen välillä tai optimaalinen funktiomuoto on vaihdellut tutkimuksesta toiseen. Selityksenä on tarjottu sitä, että otos ei ole edustava ja hyvin monet funktiomuodot pystyvät hyvään approksimaatioon paikallisesti (katso esimerkiksi [RGBP91, sivu 62]). Vaikka funktiomuoto valitaan tässäkin tutkimuksessa osin empiirisesti, on ekonometrinen mallimme selvästi erilainen tutkimusalueelle tyypillisiin Box-Cox tai logarimisiin malleihin verrattuna.

Teoreettinen tutkimus Rosenin jälkeen on valaissut hieman funktiomuodon valintaongelmaa. Jones [Jones88] on osoittanut, että markkinoilla, joilla ei ole tuotantoa[2] ja kuluttajien ollessa hinnan ottajia, hintafunktion täytyy olla konveksi ominaisuuksiensa suhteen. Lisäksi mallin kuluttajien hyöty saa riippua vain ominaisuuksien määrästä. Vain mikäli preferenssit ovat homoteetteja[3] hintafunktio voi olla lineaari. Jonesin todistuksen arvoa nostaa se, ettei se edellytä mitään kilpailumuodolta, mutta tämän vastapainona edellytetään ominaisuuksien täydellinen ja vapaa korvattavuus. Masonin ja Quigleyn [MQ96] artikkelissa kritisoidaan näkemystä, että hedoninen hintafunktio voitaisiin johtaa mikrotalousteorian abstrakteista periaatteista. Heidän näkemyksensä mukaan pelkkä teoria ei kerro käytännössä mitään hedonisen hintafunktion muodosta. Lisäksi tulevat institutionaaliset ja historialliset syyt, jotka estävät teorian toteutumisen käytännössä.

Tulevaisuutta ennustavissa malleissa voi funktiomuodon valinnan jättää osin teoreettisten perustelujen varaan [RGBP91, sivu 391] erityisesti, jos teoria toimii vain näkemyksen tukena. Tässä tutkimuksessa tärkein tavoite on asuntomarkkinoiden kuvaus, eli pyrimme mahdollisimman hyvään, mutta samalla tiiviiseen, esitykseen aineistosta. Juuri tässä suhteessa, aineiston kuvauksessa käytettävät ei-parametriset menetelmät ovat vahvimmillaan.

Oletamme, että hintafunktio on ainakin parametriensa suhteen lineaarinen. Merkitään asuntojen hintahavaintoja $(n \times 1)$ vektorilla $\mathbf{p}$ ja hintaeroja selittäviä havaintoja $(n \times K)$ havaintomatriisilla $\mathbf{H}$ sekä satunnaista virhetermiä vektorilla $\boldsymbol{\epsilon}$. Hintafunktio on

$$\mathbf{p} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

missä $\boldsymbol{\beta}$ on $(K \times 1)$ parametrivektori. Mikäli selittävien muuttujien matriisi ei sisällä mitään muunnoksia muuttujista, on hintafunktio myös asuntojen ominaisuuk-

[2] Esimerkiksi kauppa kiinteällä asuntokannalla ilman korjausrakentamista.

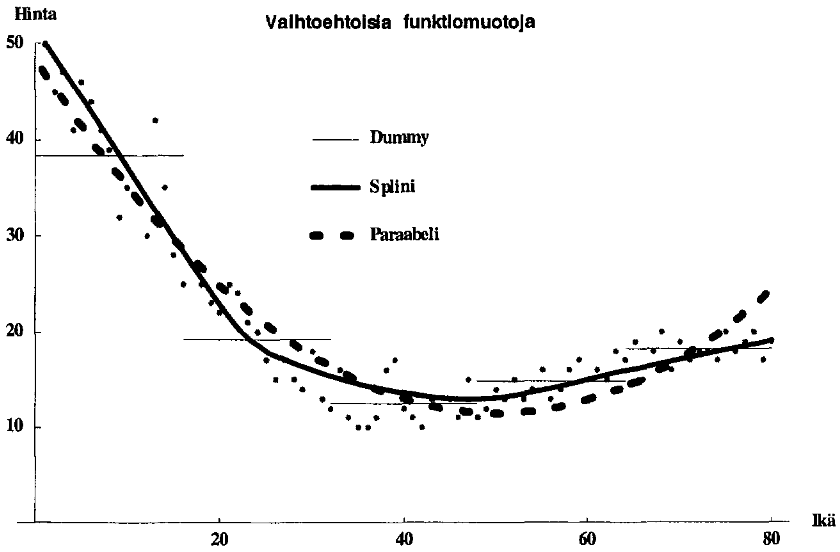
[3] Homoteetilla hyötyfunktiolla ominaisuuksien korvattavuus on riippuvainen vain niiden suhteesta, ei tasosta.

sien suhteen lineaari. Matriisi **H** voi kuitenkin sisältää muuttujasarjojen eksponenttimuunnoksia (esimerkiksi neliömuotoja varten) tai dummy-muuttujia. Kuvaustapoja (muunnoksia) on rajaton määrä. Kaikkein suurimman vapauden kuvaukselle antaa ei-parametrinen funktio. Tässä vertailemme parametrisiin kuvauksiin yhtä ei-parametrisen funktioperinteen edustajaa, *splini-funktiota(spline function, smoothing spline)* tai tutummin *ranskalaista käyrää (French curve)*. Se matemaattisiin perusteisiin palataan myöhemmin. Toteamme tässä vain että ranskalainen käyrä myötäilee kaksiulotteisessa kuviossa vapaasti havaintojen osoittamia muutoksia.

Vertailu eri kuvaustapojen välillä on helpointa esittää graafisesti. Kuviossa *1* olemme piirtäneet sovitteen kuvitteelliselle pisteparvelle dummy-mallilla (vaakasuorat viivat), neliömuodolla (katkoviiva)ja splinillä (yhtenäinen viiva). Pisteet on sijoitettu siten, että ne voisivat kuvata esimerkiksi asunnon iän ja hinnan välistä suhdetta. Hinta laskee ensin voimakkaasti, tasaantuen ja kääntyen nousuun esimerkiksi korjausrakentamisen tai rakennustyylin vuoksi. Dummy-malli antaa tyydyttävän yleiskuvan havaituista muutoksista, mutta sen perusteella ei pysty laskemaan funktion muutosnopeuksia tietynikäisille asunnoille. Neliömuotoinen polynomi sopii "havaintoihin" päällisin puolin hyvin, mutta sitä sitoo havaintoihin kuulumaton symmetrisyys; vanhojen asuntojen hintojen nousu on loivempaa kuin uusien lasku. Siten neliömuodosta laskettu sovite tuottaa systemaattisen virheen; hintojen lasku jatkuu kauemmin kuin pitäisi ja vanhimpien asuntojen hintaa yliarvioidaan. Splini-funktio on neliömuodon tapaan polynomi, mutta sen parametrit muuttuvat havaintoalueella. Tämä mahdollistaa vapaasti polveilevan kuvauksen ja hyvän sovitteen. Tällä vapaudella on oma hintansa. Samoin kuin dummy-mallissa, splinille täytyy valita joukko vaaka-akselin pisteitä, solmukohtia, joissa sen arvo ja derivaatat lasketaan. Kuvion splini-käyrä on laskettu siten, että se kulkee kuuden ennalta määrätyn pisteparin kautta. Jatkossa esitämme estimointimenetelmän, jossa sovite määritellään solmukohtien virheen ja mallin epälineaarisuuden perusteella.

Ei-parametriset menetelmät eivät toistaiseksi kuulu ekonometrikon vakio työkalupakkiin. Kiinnostus on kuitenkin heräämässä, ei ehkä vähiten Yatchewin [Yat98] surveyn vaikutuksesta. Jatkokurssin tasoiset kirjat (esimerkiksi Greene [Gre93] ja Johnston [Joh84]) kyllä tuntevat kernel-estimoinnin ja splinin, mutta eivät mene juuri esittelyä pidemmälle. Peruskurssitason kirjoille lähestymistavat ovat tuntemattomia, vaikka lineaarinen splini-regressio tekee vähemmän oletuksia aineistosta kuin kiinteäparametrinen malli. Vertailua regressiomallien ja lineaarisen splinin välillä on tehnyt ainakin Laakso [Laa97, kappale 6], jonka sovellus koskee Helsingin asuntomarkkinoita. Tässä tutkimuksessa käyttämämme kuutiomuotoon perustuvaa spliniä.

Ei-parametrisella regressiolla tarkoitetaan huomattavasti laajempaa joukkoa es-



Kuvio 1: Toisen asteen polynomi-, dymmy- ja splinifunktiosovite kuvitteellisille havaintojoukolle

timaattoreita, joka kattaa myös muuttuvaparametriset polynomit. Polynomin aste voi olla mikä tahansa, mutta jo kolmannen asteen polynomi tarjoaa riittävän epälineaarisen kuvauksen, jos annamme sen parametrien muuttua valituissa solmukohdissa. Ekonometrikon tehtävänä on löytää optimaalinen parametrisointi splinille. Valitettavasti pienimmän neliösumman kriteeri sellaisenaan sopii huonosti tähän tarkoitukseen. Estimaattori voidaan kuitenkin johtaa siten, että projektio havainnoista $\mathbf{p}$, sovitteeseen $\widehat{\mathbf{p}}$, on lineaari, eli etsimme matriisia $\mathbf{S}$ siten, että[4]

$$\widehat{\mathbf{p}} = \mathbf{S}\mathbf{p}.$$

Tähän asetelmaan palaamme seuraavan kerran vasta kappaleessa 2.6, mutta siihen liittyvä lineaarinen estimaattori rakennetaan jo kappaleessa 2.5.

Hedoninen teoria edellyttää, että hintafunktioon liitetään kaikki asunnon ominaisuudet. Splinien laskeminen useammalle muuttujalle ja niiden yhdistäminen samaan malliin on kaikkein helpointa olettamalla splini-käyrät additiivisiksi. Estimaattorin ominaisuudet monimuutujaympäristössä eivät ole kuitenkaan tarkkaan tiedossa, joten käytännöllisempää on yhdistellä parametrisia ja ei-parametrisia komponentteja malliin. Tuloksena on käyttämämme semiparametrinen funktio.

Tämän kappaleen loppu etenee kohti semiparametrisen mallin operationalisointia, eli splinin matemaattisesta hallinnasta sen laskenta-algoritmeihin ja lopuksi

[4]Vertaa esimerkiksi PNS regression projektioon[Gre93, sivu 244].

tilastolliseen päättelyyn. Kappaleessa 2.2 selvitämme millainen funktio tarvitaan ranskalaisen käyrän piirtämiseen ja kappaleessa 2.3 teemme katsauksen sen estimointimenetelmään. Seuraavassa kappaleessa 2.4 laajennamme keskustelun semiparametriseen malliin ja sen estimointiperiaatteeseen. Loput tämän kappaleen alaluvuista käsittelevät semiparametrisen mallin estimaattoria ja sen operationalisointia laskenta-algoritmissa. Erityisesti kappaleissa 2.6 ja 2.7 esitämme menetelmiä mallin ja estimaattien stabiilisuuden arviointiin. Menetelmävalintamme pyrkivät tuomaan esille estimointimenetelmän perusteita, ei kaikkia siihen sisältyviä mahdollisuuksia.

2.2 Splini-käyrän matemaattinen rakenne

Rajaamme keskustelua alussa siten, että selitämme asunnon hintaa p vain yhdellä hedonisella muuttujalla h. Sen sijaan, että parametrisoisimme hintafunktiota $p(h)$ tarkemmin, oletamme sen polynomiksi.

Perinteisesti ekonometrian oppikirjoissa estimoidaan jonkin polynomin $p(h)$ parametrit siten, että virhetermin neliösumma tai joku muu uskottavuuskriteeri täyttyy. Lähestymistapa ei ole täysin realistinen, mutta käytännöllinen, sillä parametriestimaatit tiivistävät tietoa tavalla, joka on annetuin oletuksin mallin stokastiikasta helposti tilastollisesti testattavissa. Havaintoaineiston kuvaus on kuitenkin riippuvainen valitusta polynomista. Sen käyttäytymistä koko määrittelyalueellaan dominoi havaintoalue tai pahimmassa tapauksessa jonkin havaintoalueen osan vaihtelu. Havaintojen kuvausta voidaan parantaa dummy-muuttujilla tai sallimalla parametrien muutokset eri havaintoalueen osilla. Valintoja voidaan ohjata tilastollisilla testeillä, mutta tuloksena on epäjatkuva kuvaus havaintoaineistosta ja polynomien uudelleen parametrisointi jokaiselle havaintovälille maksaa vapausasteita. Tässä tutkimuksessa on haluttu antaa jatkuva, mutta mahdollisimman tarkka kuvaus aineistosta.

Määritelmä[5]: *Olkoot a ja b splini-funktion määrittelyalueen rajat (havaintovälin ulkopuolelta) ja pisteet t_i, $a = t_0 < t_1 <, \dots, < t_M < t_{M+1} = b$ valittuja solmukohtia. Estimoitava hintafunktio, $p(h)$, on splini, jos se on*

- *korkeintaan n asteinen muuttujan h:n polynomi jokaisella välillä $[t_{i-1}, t_i]$, $1 \leq i \leq M+1$.*
- *Sen $(n-1)$ ensimmäistä derivaattaa ovat jatkuvia välillä $[a, b]$.*

Merkitään splini-funktiota jatkossa $P(\cdot)$. Funktion ääripäitä a ja b ei myöhemmin tehtävien rajausten perusteella tarvitse määritellä numeerisesti. Käytännössä niitä

[5]Tämä määritelmän takana on Khuri [Khu93, sivu 422]. Splini-funktion on ensimmäisenä esittänyt Schoenberg [Sch46a] ja [Sch46b]

tarvitaan vain varmistamaan derivaattojen jatkuvuus havaintoalueen ensimmäisessä ja viimeisessä solmussa. Määritelmän mukaisesti esimerkiksi jaksoittain lineaari käyrä, missä suorat liittyvät toisiinsa solmukohdissa on splini-funktio. Kuitenkin jo toisen asteen splinillä on solmukohtien kaarevuus sama riippumatta siitä, mistä suunnasta solmua lähestytään, eikä silmämääräinen tarkastelu välttämättä paljasta solmukohtia lasketusta splinistä, vaikka polynomin kertoimet siinä muuttuvatkin määrätyllä tavalla.

Splini-funktio voidaan esittää yleisemmin muodossa

$$P(h) = \sum_{i=1}^{m} e_i \, (h - t_i)_+^r + g(h), \text{ missä} \tag{2}$$

$$(h - t_i)_+^r = \begin{cases} (h - t_i)^r & h \geq t_i \\ 0 & h \leq t_i \end{cases}, \; g(h) \text{ polynomi, } e_i \text{ vakio} \tag{3}$$

Funktio (3) tunnetaan myös nimellä Greenin funktio. Jatkossa merkitsemme jotain splinin osaa funktiolla $g(\cdot)$, mahdollinen alaindeksi viittaa solmuväliin[6].

Oletetaan, että $g(h)$ yhtälössä (2) on kuutiomuotoinen ($r = 3$). Solmukohdan alapuolella ($h \leq t$) splini on $a_1 + b_1 h + c_1 h^2 + d_1 h^3$. Siten parametrien muunnos solmukohdan yläpuolelle tapahtuu (2):n mukaan

$$a_2 + b_2 h + c_2 h^2 + d_2 h^3 = e\,(h - t)^3 + a_1 + b_1 h + c_1 h^2 + d_1 h^3$$

Kun avaamme kuutiotermin $(h - t)^3$ ja vertaamme keskenään samanasteisia h:n termejä, saadaan

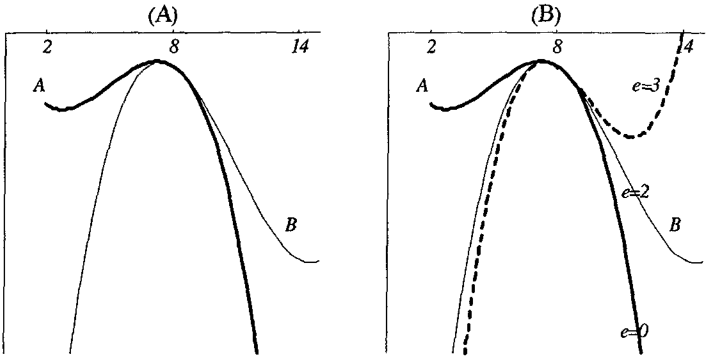
$$a_2 + b_2 h + c_2 h^2 + d_2 h^3 = e\left(h^3 + 3ht^2 - 3h^2 t - t^3\right) + a_1 + b_1 h + c_1 h^2 + d_1 h^3$$

$$\Longrightarrow \begin{array}{rcl} a_2 & = & a_1 - et^3 \\ b_2 & = & b_1 + 3et^2 \\ c_2 & = & c_1 - 3et \\ d_2 & = & d_1 + e \end{array} \tag{4}$$

Tästä nähdään, että splinin uudelleen parametrisointi vaatii vain yhden lisäparametrin estimointia per solmukohta. Mikäli jokaiselle solmuvälille on käytettävissä runsaasti havaintoja, voidaan havaintomatriisia sopivasti järjestelemällä (katso esimerkiksi Johnston, [Joh84, sivut 392-396]) estimoida splinin polynomit jokaiselle solmuvälille PNS menetelmällä.

Splinin muodostamista on nyt helppo simuloida numeerisesti. Oletetaan, että ensimmäisen solmuvälin polynomi on muotoa $-60h + 15h^2 - h^3$ ja $e = 2$. Valitaan solmukohdaksi $h = 8 = t$. Splinin toinen osa on $(0 - 2 \times 8^3) + (-60 + 3 \times 2 \times 8^2)\,h + (15 - 3 \times 2 \times 8)\,h^2 + (-1 + 2)\,h^3$. Kuviossa *2* (A) on piirretty molemmat käyrät erikseen. Ohuempi käyrä määrittelee splinin välillä $[8, \infty]$.Koko splini on kuvaaja AB.

[6] Tämä määritelmä johtaa samoihin parametrirajoitteisiin kuin liitteessä 1, mutta havainnollistaa vapausasteiden kulumista paremmin.



Kuvio 2: Splini-käyrän jatkuminen solmupisteessä

Kuviossa 2 (B) on piirretty splinejä vaihtoehtoisilla parametrin e arvoilla $(0, 2, 3)$.

Splinin rakentaminen edellyttää polynomin parametrien identifiointia jokaiselle solmuvälille. Tämä aiheuttaa ilmeisen identifiointiongelman varsinkin, jos solmuvälillä ei ole tai on vain vähän havaintoja. Tarkastellaan tilannetta siten, että kaikki käytettävä informaatio on solmukohdissa $a = t_0 < t_1 <, \ldots, < t_M < t_{M+1} = b$. Eli solmuja on $M + 2$ kappaletta ja solmuvälejä $M + 1$. Indeksoidaan välit niiden alarajan mukaisesti, eli solmuväleillä $i = 0, \ldots, M$ polynomeja merkitään $g_i(h)$. Jos estimoimme kuutiopolynomin jokaiselle välille, on estimoitavia parametrejä $4(M + 1)$. Parametritransformaatioiden (4) perusteella tarvitaan vain yksi uusi parametri e jokaiselle välille, kunhan neljä parametriä on ensin annettu jollekin välille eli vapaita parametrejä on vain $M + 4$ kappaletta. Koska $M + 2$ solmusta (yhtälöstä) ei kuitenkaan voida ratkaista $M + 4$ parametriä, malli on ali-identifioitu. Splini-funktion parametrit identifioidaan luomalla kaksi edellisistä riippumatonta parametrirajoitetta. Vähiten harmia tuotetaan asettamalla rajoitteet havaintoja rajoittaviin solmuväleihin $[a, t_1]$ ja $[t_M, b]$ siten, että

$$\begin{aligned} g_0''(t_1) &= \pi_a g_0''(a), \\ g_M''(t_M) &= \pi_b g_M''(t_{M+1}), \end{aligned}$$

missä π_a ja π_b ovat tunnetut vakiot. Rajoitteen asettamisen jälkeen spliniin jää $M + 2$ vapaata parametriä, jotka voidaan estimoida vastaavasta määrästä solmukohtia. Luonnolliseksi kuutiosplini *(natural cubic spline)* saadaan asettamalla vakiot nollaksi. Tästä seuraa, että

$$\begin{array}{rcccl} g_0''(t_1) &=& \gamma_1 &=& 0, \\ g_M''(t_M) &=& \gamma_M &=& 0. \end{array} \tag{5}$$

Derivaattojen jatkuvuusehto solmuissa takaa myös sen, että luonnollinen splini muuttuu lineaariseksi lähestyttiinpä äärimmäisiä havaintoja kummalta puolen tahansa.

Luonnollisen splinin määrittelyyn *havaintoalueen* solmukohdissa tarvitaan splinin arvot jokaisessa havaintoalueen M solmussa, mutta vapaita toisia derivaattoja on vain $M-2$ kappaletta, sillä $\gamma_1 = \gamma_M = 0$. Merkitään näitä valintoja jatkossa M vektorilla $\mathbf{g} = [g_1(t_1), \ldots, g_M(t_1)]^{\top} = [g_1, \ldots, g_M]^{\top}$ja $M-2$ vektorilla $\gamma = [\gamma_2, \ldots, \gamma_{M-1}]^{\top}$.

2.3 Splini-käyrän estimointikriteeri

Splini-käyrä voidaan estimoida pienimmän neliösumman periaatteella, mutta polynomin hyvä mukautumiskyky tekee sovitteesta täydellisen eli tulokseksi saadaan hyvin epälineaari havainnosta havaintoon aaltoileva kuvaaja. Koska tämä ei mitenkään tiivistä informaatiota havaintoaineistossa, käyrää yritetään tasoittaa rankaisemalla sitä epälineaarisuudesta. Merkitään hintahavaintoja solmupisteissä p_i. Estimointikriteeri $S(P(h))$ on muotoa

$$S(P(h)) = \sum_i (p_i - g_i)^2 + \alpha J(P(h)), \alpha \geq 0, \tag{6}$$

missä $J(\cdot)$ on rankaisutermi ja α sen paino. Rankaisutermille ei voida määrittää yleisesti pätevää optimaalista muotoa vain tehtäväasettelun mukaiset reunaehdot. Ensikädessä sen tulee vaikuttaa estimointikriteeriin vain jos $P(h)$ on epälineaari. Yksinkertainen ratkaisu tähän on käyttää splinin toista derivaattaa $P''(h)$, mistä mahdolliset vakiotermit ja lineaarinen termi ovat derivoituneet pois. Sellaisenaan toinen derivaatta sopii huonosti rankaisutermiksi, sillä sen arvo voi vaihdella polynomin asteen ja havaintojen arvojen mukaisesti. Sen sijaan estimointikriteeriin tulisi liittyä rankaisu epälineaarisuuden "koosta" annetulla havaintoaineiston osalla. Toisen derivaatan normi käy mittariksi, sillä funktion f normille $\|f\|$ [7]

$$\|f\| > 0 \text{ joss } f \neq 0, \tag{7}$$

$$\|af\| = |a| \, \|f\| \, , a \in \Re \tag{8}$$

$$\|f + g\| \leq \|f\| + \|g\| \, . \tag{9}$$

Eli normi on aina positiivinen, funktion skalaarilla kertominen transponoi normia skalaarin itseisarvolla ja kahden normin summa ei ole pienempi kuin summan normi. Eräs normin ominaisuudet täyttävä funktio on kaikki polynomit sisältävän ja kahdesti derivoituvien funktioiden avaruuden $L_2[a, b]$ normi

$$L_2(f(h)) = \|f(h)\|_2 = \left(\int_a^b [f(h)]^2 \, dh \right)^{0.5} .$$

[7] Selkeä mutta tiivis esitys funktioavaruuden normeista on esimerkiksi Pearsonilla [Pea83].

Tämän normin minimoiva funktio on

$$\min_f \|f\|_2 = \min_f \int_a^b [f(h)]^2 \, dh. \tag{10}$$

Haetaan lausekkeen minimiä siten, että $f = P''(h)$. Sijoitetaan ja käytetään sitä rankaisuterminä eli

$$J(P(h)) = \int_a^b [P''(h)]^2 \, dh. \tag{11}$$

Funktio $P''(h)$ kuvaa funktion muutosnopeutta havaintovälillä. Sen neliöinti poistaa etumerkin vaikutuksen ja integraali "summaa" nopeuden muutokset määrittelyalueelta. Se on nolla vain, jos splini-funktio on lineaari koko määrittelyalueellaan, muutoin se on nollaa suurempi. Jatkossa kirjoitamme splinin estimointikriteerin (6) siten, että $P(h)$ minimoi summan $\sum (p_i - g_i)^2 + \alpha \int_a^b [P''(h)]^2 \, dh$.

Rankaisutermi on hyvin epälineaarinen funktio, mutta sillä voidaan osoittaa olevan yksikäsitteinen minimiratkaisu kahdesti derivoituvien funktioiden joukossa. (Todistus esimerkiksi Eubank [Eub88]). Kovin yllättävä tulos ei tosin ole se, että minimiratkaisu on kuutiopolynomi. Korkeamman asteen polynomit eivät sisällä mitään käyrän suunnanmuutosten kannalta relevanttia lisätietoa solmuvälillä. Niinpä kuutiopolynomin valinta splinin pohjaksi estimointialgoritmissä ei ole optimaalisuuden kannalta tarkastelua rajoittava tekijä. Samalla se yksinkertaistaa merkittävästi rankaisutermiä, sillä sen toinen derivaatta on lineaarinen ja kolmas derivaatta vakio. Osittaisintegroimalla[8] (11) saadaan

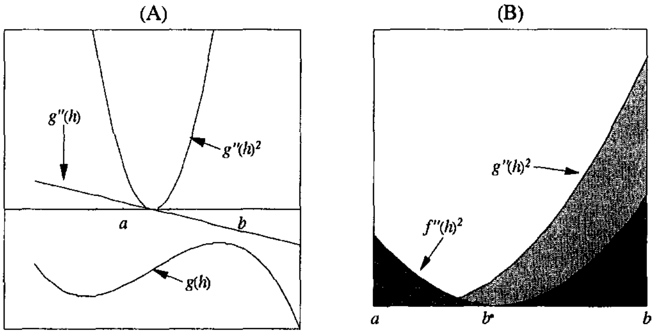
$$\int_a^b [P''(h)]^2 \, dh = g''(b)g'(b) - g''(a)g'(a) - \int_a^b P'''(h)P'(h)dh.$$

Merkitään vakiota $g''(b)g'(b) - g''(a)g'(a) = \xi$.

$$\begin{aligned}
\int_a^b [P''(h)]^2 \, dh &= \xi - \int_a^b P'''(h)P'(h)dh. \\
&= \xi - P'''(h) \int_a^b P'(h)dh \\
&= \xi - \sum_{i=0}^{M} g_i'''(t_i) \int_{t_i}^{t_{i+1}} g_i'(h)dh \\
&= \xi + \sum_{i=0}^{M} g_i'''(t_i) \left[g_i - g_{i+1}\right].
\end{aligned} \tag{12}$$

Välimuodot saadaan ottamalla huomioon kuutiopolynomin vakio kolmas derivaatta ja hajottamalla integraali summaksi yli solmuvälien. Viimeisessä muunnoksessa integroitava funktio on derivaatta eli $\int P'(h)dh = P(h)$. Jatkossa teemme lisäoletuksen, joka tekee ξ:n nollaksi, joten tarkastellaan vain oikeanpuoleista termiä.

[8]Merkittään $u = g''(h)$ ja $v = g'(h)$, sijoitetaan osittaisintegrointikaavaan $\int uv'dx = uv - \int u'vdx$.



Kuvio 3: Rankaisutermin graafinen tulkinta

Kuutiosplinin tapauksessa splinin muutosta painotetaan vakiolla. Koko termi on aina positiivinen (yhtälöiden (7) ja (10) nojalla) eli (6):n optimoinnissa pyritään eliminoimaan splinin voimakkaat muutokset, erityisesti toisen derivaatan muutokset. Kuitenkin ainoa vakio, minkä toinen derivaatta voi saavuttaa, on nolla. Tämä nähdään parhaiten seuraavassa graafisessa tarkastelussa.

Tarkastellaan polynomia ja rankaisutermiä määrittelyalueella $[a, b]$. Ilman, että tarkastelun yleisyys siitä mitenkään kärsii, oletetaan splinin koostuvan vain yhdestä osasta,

$$g(h) = a_i + b_i h + c_i h^2 + d_i h^3 \qquad g''(h) = 2c_i + 6d_i h,$$
$$[g''(h)]^2 = 4c_i^2 + 24c_i d_i h + 36\,(d_i h)^2 \quad \int [g''(h)]^2\, dh = 4c_i^2 h + 12 c_i d_i h^2 + 12 d_i^2 h^3$$

Koska kuutiopolynomille $g''(h)$ on lineaari, integroitava lauseke on aina neliömuoto. Voidaan myös osoittaa, että se on ylöspäin aukeava paraabeli ja sillä on kertaantuvat juuret. Näin sen ja vaaka-akselin rajoittama pinta-ala välillä $[a, b]$ (eli rankaisutermi) on aina positiivinen ja kasvaa jos tarkasteluväliä kasvatetaan. Kuviossa *3* (A) on piirretty polynomi ja sen derivaatat. Toista derivaattaa kuvaava suora voi olla vakio vain arvolla nolla, eli rankaisutermi pyrkii linearisoimaan splinin. Rankaisutermi on kuviossa *3* (B) paraabelien alle jäävä alue välillä $[a, b]$ ja sen pinta-ala minimoituu levittämällä paraabelia (tai vastaavasti tekemällä $g''(h)$ horisontaaliseksi). Tämä onnistuu vain asettamalla $c_i = d_i = 0$ eli splini linearisoituu.

Kuten integraalilausekkeen (12) analyysistä jo nähtiin, linearisoiva vaikutus ei kuitenkaan ole riippuvainen vain derivaatoista. Kuviossa *3* (B) vertaillaan kahta vaihtoehtoista parametrisaatiota solmuvälille $g''(h)^2$ ja $f''(h)^2$ siten, että[9] $c_f \neq c_g$ ja $d_f = d_g$. Jo silmämääräinen tarkastelu paljastaa, että funktion $[g''(h)]^2$ rajaama pinta-ala välillä $[a, b]$ –samalla rankaisutermi– on suurempi, vaikka rankaisutermin

[9]Korvataan yleensä solmuväliä kuvaava parametrin alaindeksi funktiota kuvaavalla indeksillä.

painona oleva kolmas derivaatta on molemmissa sama. Nyt $[g_a - g_b] > [f_a - f_b]$ eli muutokset funktion $g(h)$ arvossa ovat voimakkaampia ja algoritmi pyrkii valitsemaan $f(h)$:n.

2.4 Additiivinen ja semiparametrinen malli

Tähän asti käsitelty ei-parametrinen malli on käsittänyt vain yhden ei-parametrisen muuttujan, jonka vaihtelua on kuvattu splini-funktiolla. Usean muuttujan mallissa joudumme tyytymään additiiviseen malliin

$$p = \sum_{j=1}^{K} P_j(h^j) + \epsilon, \tag{13}$$

missä indeksi j kulkee yli muuttujien. Kuten lineaarinenkin malli (1), ei tämäkään täytä hedonisen hintafunktion vaatimaa muuttujien interaktiota, mutta nyt summan elementtien ominaisuuksia ei ole rajattu. Osa muuttujista voi mallissa vaikuttaa hintaan lineaarisestikin. Erotetaan näiden muuttujien havainnot omaksi $(n \times k)$ -matriisiksi $\mathbf{X}$. Yhdistämällä parametriset (lineaariset) muuttujat additiiviseen malliin saadaan semiparametrinen regressiomalli

$$\mathbf{p} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^{l} \mathbf{g}_j + \epsilon,$$

missä $\mathbf{g}_j$ on estimaattivektori j:nnen muuttujan tasosta ja $k + l = K$. Tämä on tarkkaan ottaen ristiriitainen tapa kirjoittaa semiparametrinen malli. $\mathbf{p}$ ja $\mathbf{X}\boldsymbol{\beta}$ ovat $(n \times 1)$ vektoreita, mutta splinikäyrän korkeus estimoidaan vain järjestelyille ja eikertaantuville havainnoille. Toisin sanoen solmujen määrä on käytännössä aina alle havaintojen ja ne ovat toisessa järjestyksessä kuin muiden muutujien havainnot. Jos $\mathbf{g}_j$:n dimensio on M_j, ongelma voidaan kiertää muuttujakohtaisella $(n \times M_j)$ järjestysmatriisilla *(incident matrix)* $\mathbf{N}_j$, joka kuvaa havaintojen esiintymistiheyttä ja järjestystä. Järjestetyt arvot voidaan palauttaa havaintoja vastaavaan järjestykseen operaatiolla $\mathbf{N}_j\mathbf{g}_j$. Esimerkki selventäköön matriisin rakennetta.

Olkoon $\mathbf{g}_j = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ ja $\mathbf{N}_j = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. Siten havainnot ovat $\begin{bmatrix} 3 \\ 4 \\ 3 \\ 2 \end{bmatrix}$.

Riippumatta siitä, miten monta kertaantuvaa elementtiä kussakin muuttujassa on ja miten ne ovat järjestetty, semiparametrinen malli voidaan aina esittää muodossa

$$\mathbf{p} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^{l} \mathbf{N}_j\mathbf{g}_j + \epsilon. \tag{14}$$

Lineaarisen mallin ja splinin yhdistäminen edellyttää yksittäisen splinin estimointikriteeriin vain pieniä muutoksia. Se voidaan johtaa sijoittamalla yhtälö (14) ja rankaisutermi (11) estimointikriteeriin (6).

$$\min_{\beta,g_j} S(\beta,g_j) = (\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^{l} \mathbf{N}_j \mathbf{g}_j)^{\mathsf{T}} (\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^{l} \mathbf{N}_j \mathbf{g}_j) +$$
$$\sum_{j=1}^{l} \alpha_j \int_a^b \left[P_j''(h^j)\right]^2 dh^j. \tag{15}$$

Ei-parametrinen (6) ja semiparametrinen estimointikriteeri (15) poikkeavat toisistaan solmukohtien määrittelyssä. Yhden muuttujan ei-parametrisessa mallissa voidaan solmujen sijainti ja lukumäärä valita vapaasti. Semiparametrisessa mallissa jokainen ei-kertaantuva havainto on solmu. Jos ei-parametriset havainnot saavat arvoja vain hyvin rajatulla alueella, kuten aikatrendi, laskennalliset ongelmat helpottuvat huomattavasti.

2.5 Semiparametrisen mallin estimointi

2.5.1 Splinin parametrirajoitteet

Edellä olemme esittäneet splini-polynomin parametrirajoitteet yhdessä solmussa. Seuraavaksi esitämme vastaavat rajoitteet jokaisessa solmussa splinin ja sen toisen derivaatan arvon avulla. Esitystapa perustuu pitkälti Wegmanin ja Wrightin artikkeliin [WW83]. Tarkastellaan aluksi semiparametrisen mallin estimointia yhdellä ei-parametrisella muuttujalla h. Merkitään sen järjestettyjä havaintosolmuja $[t_1, \ldots, t_M]$. Kirjoitamme kuutiopolynomin (jokaisen splinin osan) poikkeamana solmukohdasta t_i, eli[10]

$$g_i(h) = D_i(h - t_i)^3 + C_i(h - t_i)^2 + B_i(h - t_i) + A_i, \tag{16}$$

Tarkasteltaessa polynomia $g_i(h)$ solmukohdissa t_i ja t_{i+1}, määrittelyalueensa ylä- ja alapäässä, saadaan sen arvoiksi

$$g_i(t_i) = A_i \tag{17}$$
$$g_i(t_{i+1}) = D_i(t_{i+1} - t_i)^3 + C_i(t_{i+1} - t_i)^2 + B_i(t_{i+1} - t_i) + A_i$$
$$= D_i T_i^3 + C_i T_i^2 + B_i T_i + A_i \tag{18}$$

[10]Polynomi $a_i + b_i h + c_i h^2 + d_i h^3$ (katso myös liitettä 1) voidaan johtaa poikkeamamuodon parametreistä muunnoksin $d_i = D_i$, $c_i = (C_i - 3D_i t_i)$, $b_i = B_i. - 2C_i t_i + 3D_i t_i^2$, $a_i = A_i - B_i t_i + C_i t_i^2 - D_i t_i^3$

eli yksinkertaistetaan lausekkeita merkitsemällä solmuvälin erotusta $t_{i+1} - t_i = T_i$.

Sen sijaan, että kirjoitetaan spliniä rajoittavat jatkuvuusehdot (4) ja luonnollisen splinin asettama parametrirajoite uudella notaatiolla, muokataan nämä ehdot sisältävä lauseke jokaiselle havaintosolmulle[11] $[t_1, t_M]$. Polynomin (16) toinen derivaatta $g_i''(h) = 6D_i(h - t_i) + 2C_i$ saa solmukohdassa t_i, arvon $g_i''(t_i) = \gamma_i = 2C_i$, eli

$$C_i = \frac{\gamma_i}{2}. \tag{19}$$

Funktion arvon sekä ensimmäisen että toisen derivaatan tulee olla jatkuvia kaikissa solmukohdissa. Asettamalla toiset derivaatat $g_i''(h)$ ja $g_{i+1}''(h)$ yhtäsuuriksi solmukohdassa t_{i+1} saadaan

$$\begin{aligned} 6D_iT_i + 2C_i &= 2C_{i+1} \\ 6D_iT_i + \gamma_i &= \gamma_{i+1} \\ D_i &= \frac{(\gamma_{i+1} - \gamma_i)}{6T_i}. \end{aligned} \tag{20}$$

Polynomin jatkuvuuden nojalla solmussa t_{i+1} pätee $g_i(t_{i+1}) = g_{i+1}(t_{i+1})$. Siten yhtälöitä (17) ja (18) avulla voidaan kirjoittaa

$$D_iT_i^3 + C_iT_i^2 + B_iT_i + A_i = A_{i+1}.$$

Sijoittamalla tähän(17), (19) ja (20) saadaan ratkaistua parametri B_i,

$$B_i = \frac{A_{i+1} - A_i}{T_i} - \frac{D_iT_i^3 + C_iT_i^2}{T_i} = \frac{g_{i+1} - g_i}{T_i} - \frac{(\gamma_{i+1} + 2\gamma_i)T_i}{6}. \tag{21}$$

Ensimmäisten derivaattojen jatkuvuus solmukohdissa, $g_i'(t_{i+1}) = g_{i+1}'(t_{i+1})$ täydentää parametrirajoiteet, eli

$$3D_iT^2 + 2C_iT + B_i = B_{i+1}.$$

Jo nyt nähdään, että splinin parametrien ratkaisu riippuu vektoreista $\mathbf{g}$ ja $\boldsymbol{\gamma}$. Sijoittamalla muista jatkuvuusehdoista lasketut parametrit saadaan lauseke, joka toteutuessaan määrittää kuutiosplinin solmussa t_i,

$$\frac{1}{6}T_{i-1}\gamma_{i-1} + \frac{1}{3}(T_{i-1} + T_i)\gamma_i + \frac{1}{6}T_i\gamma_{i+1} = \frac{g_{i+1} - g_i}{T_i} - \frac{g_i - g_{i-1}}{T_{i-1}} \tag{22}$$

[11] Emme ole kiinnostuneita havaintovälin ulkopuolella splinin parametreistä.

Huomionarvoista tässä on, että oikealla puolella on splinin derivaattojen erotus perättäisissä solmuissa ja vasemmalla puolella solmuväleillä painotettu summa toisista derivaatoista. Kuten rajoiteryhmä (4) on tämäkin yhtälö lineaarinen. Harjoitelma tuottaa vektorimuodon,

$$\left[\frac{1}{6}T_{i-1}, \frac{1}{3}(T_{i-1}+T_i), \frac{1}{6}T_i\right]\left[\gamma_{i-1}, \gamma_i, \gamma_{i+1}\right]^\top = \left[T_{i-1}^{-1}, -T_i^{-1} - T_{i-1}^{-1}, T_i\right]\left[g_{i-1}, g_i, g_{i+1}\right]^\top. \tag{23}$$

Tämä yhtälö ei vielä ratkaise splinin parametrejä, mutta jokainen ratkaisu (estimaattori) täytyy laatia siten, että yhtälö toteutuu. Parametrien täysi identifioituvuus vaati vielä luonnollista spliniä (rajoite (5) eli $\gamma_1 = \gamma_M = 0$), joka supistaa vapaiden toisten derivaattojen γ-vektoria kahdella elementillä ja luonnollisesti niitä vastaavat kertoimet yhtälössä 22 voidaan asettaa nollaksi. Ottamalla huomioon nämä rajoitteet täytyy ensimmäiselle ja viimeiselle havaintosolmulle päteä

$$\frac{1}{3}(T_1+T_2)\gamma_2 + \frac{1}{6}T_2\gamma_3 = \frac{g_3-g_2}{T_2} - \frac{g_2-g_1}{T_1}, \tag{24}$$

$$\frac{1}{6}T_{M-2}\gamma_{M-2} + \frac{1}{3}(T_{M-2}+T_{M-1})\gamma_{M-1} = \frac{g_M-g_{M-1}}{T_{M-1}} - \frac{g_{M-1}-g_{M-2}}{T_{M-2}}. \tag{25}$$

Yhdessä (22), (24) ja (25) voidaan kirjoittaa $(M-2)$:n yhtälön ryhmänä

$$\mathbf{R}\gamma = \mathbf{Q}^\top \mathbf{g}, \tag{26}$$

joka toteutaa luonnollisen kuutiosplinin. $\mathbf{R}$ on symmetrinen $(M-2\times M-2)$ matriisi

$$\mathbf{R} = \begin{bmatrix} \frac{1}{3}(T_1+T_2) & \frac{1}{6}T_2 & 0 & \cdots & 0 \\ \frac{1}{6}T_2 & \frac{1}{3}(T_2+T_3) & \frac{1}{6}T_3 & \ddots & \vdots \\ 0 & \frac{1}{6}T_3 & \frac{1}{3}(T_3+T_4) & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \frac{1}{6}T_{M-2} \\ 0 & \cdots & 0 & \frac{1}{6}T_{M-2} & \frac{1}{3}(T_{M-2}+T_{M-1}) \end{bmatrix}$$

ja $\mathbf{Q}$ on $(M \times M-2)$ matriisi

$$\mathbf{Q} = \begin{bmatrix} T_1^{-1} & 0 & 0 & \cdots & 0 \\ -T_1^{-1}-T_2^{-1} & T_2^{-1} & 0 & \cdots & \vdots \\ T_2^{-1} & -T_2^{-1}-T_3^{-1} & T_3^{-1} & \ddots & \vdots \\ 0 & T_3^{-1} & -T_3^{-1}-T_4^{-1} & \ddots & 0 \\ \vdots & 0 & T_4^{-1} & \ddots & T_{M-2}^{-1} \\ \vdots & \ddots & \ddots & \ddots & -T_{M-1}^{-1}-T_{M-2}^{-1} \\ 0 & \cdots & \cdots & \cdots & T_{M-1}^{-1} \end{bmatrix}$$

Lineaarinen yhtälöryhmä (26) toimii optimoinnin rajoitteena estimaattorissa (15). Ratkaisussa voidaan soveltaa standardeja optimointimenetelmiä, kuten Lagrange tai rajoitteen sijoittamista suoraan optimiehtoon. Matriisien $\mathbf{Q}$ ja $\mathbf{R}$ nauhamaisuus paljastaa splinin määräämisessä rakenteen: jokaisen solmun arvot (g_i ja γ_i) määräytyvät simultaanisesti viereisten solmujen kanssa.

Verrattuna parametrisiin menetelmiin, ei-parametrisen menetelmän tulokset stabiloituvat hyvin hitaasti otoskoon kasvaessa. Käytännössä havaintoja pitää olla satoja tai tuhansia. Toinen laskennan kannalta merkittävä tekijä on solmujen lukumäärä, joka näkyy suoraan matriisien $\mathbf{Q}$ ja $\mathbf{R}$ dimensioissa. Lineaaristen yhtälöryhmien ratkaisu edellyttää poikkeuksetta kerroinmatriisien käänteismatriiseja, tässä tapauksessa $\mathbf{R}^{-1}$:n laskentaa. Jos solmujen lukumäärä kasvaa, nousee tehtävä helposti yli työpöytäkoneen laskentakapasiteetin. $\mathbf{Q}$ ja $\mathbf{R}$ ovat kuitenkin harvoja (paljon nollia) ja nollasta poikkeavat elementit sijaitsevat kolmella lävistäjällä, joten laskenta-algoritmeja taiten käyttämällä suurikin aineisto on käsiteltävissä.

Matriisin $\mathbf{R}$ rakenne on varsin yksinkertainen ja voidaan osoittaa, että se on positiivisesti definiitti, joten (katso esimerkiksi Green ja Silverman, [GS94, Sivu 13]) matriisi $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}$on olemassa. Jos tunnemme funktion arvot solmupisteissä ja rajoitematriisit, niin voimme aina laskea niitä vastaavat jatkuvat derivaatat solmuissa ratkaisemalla (26):stä)

$$\gamma = \mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{g}. \tag{27}$$

2.5.2 Rankaisutermi

Yhtälössä (12) osoitetettiin, että rankaisutermi on muotoa $\xi + \sum_{i=0}^{M} g_i'''(t_i)\,[g_i - g_{i+1}]$. Lauseke ei ole mitenkään riippuvainen splinin ominaisuuksista ja sen määrittelyalue on $[a, b]$. Mikäli splini-funktio jatkaa äärimmäisillä solmuväleillä $[a, t_1]$ ja $[t_m, b]$ täysiasteisena, ääripisteiden valinta vaikuttaisi rankaisutermiin. Koska rankaisutermi on splinin derivaattojen funktio, sitä voidaan edelleen yksinkertaistaa sijoittamalla luonnollisen splinin rajoitteet ja yhdellä tärkeällä lisäoletuksella — splini jatkaa lineaarisena havaintovälin ulkopuolella[12].

Jos splini jatkaa lineaarisena havaintovälin ulkopuolella, myös $\gamma_a = g''(a) = 0$. Vastaavasti splinin yläpäässä, niinpä vakio $\xi = g''(b)g'(b) - g''(a)g'(a) = 0$. Myös summatermissä kaikki toiset ja kolmannet derivaatat solmuissa ($t_0 = 0, t_1, \ldots, t_M$,

[12] Tässä kohdassa lähdekirjallisuus on ristiriitaista, mutta olemme valinneet käsitteet splinien käyttöä regressio-mallissa käsittelevistä lähteistä. Green ja Silverman [GS94, sivu 12] määrittävät luonnollinen splinin siten, että käyrä jatkaa lineaarisena havaintovälin ulkopuolella. Kuitenkin Poirier [Poirier73, sinu 516] sekä Buse ja Lim [BL77, sivu 64] käyttävät jo edellä esillä ollutta määritelmää jossa splini on luonnollinen jos se on lineaarinen äärimmäisissä havaintosolmuissa.

$t_{M+1} = b$) ovat nollia. Rakaisutermi supistuu näillä oletuksilla muotoon

$$\int_a^b [P''(h)]^2 \, dh = \sum_{i=2}^{M-1} g_i'''(t_i) \left(g_i - g_{i+1} \right).$$

Tavoitteena on kirjoittaa tämä vain γ_i:n ja g_i:n funktiona, joten sijoitetaan kolmas derivaatta toisen derivaatan suhteellisena muutoksena,

$$\int_a^b [P''(h)]^2 \, dh = \sum_{i=2}^{M-1} \frac{\gamma_{i+1} - \gamma_i}{T_i} \left(g_i - g_{i+1} \right).$$

Järjestelemällä termit uudelleen rankaisutermi voidaan kirjoittaa

$$\int_a^b [P''(h)]^2 \, dh = \sum_{i=2}^{M-1} \gamma_i \left(\frac{g_{i+1} - g_i}{T_i} - \frac{g_i - g_{i-1}}{T_{i-1}} \right).$$

Sulkeissa oleva lauseke on vektorin $\mathbf{Q}^{\mathsf{T}}\mathbf{g}$ elementti ja γ_i vektorin $\boldsymbol{\gamma}$ elementti, joten rankaisutermi kirjoittautuu tehdyillä oletuksilla neliömuodoksi

$$\begin{aligned} \int_a^b [P''(h)]^2 \, dh &= \boldsymbol{\gamma}^{\mathsf{T}}\mathbf{Q}^{\mathsf{T}}\mathbf{g} \\ &= \boldsymbol{\gamma}^{\mathsf{T}}\mathbf{R}\boldsymbol{\gamma}. \end{aligned} \tag{28}$$

Tämä muoto on täysin riipumaton splinin jatkuvuusehdoista. Derivaattojen jatkuvuuden vallitessa saadaan sijoittamalla (27)

$$\int_a^b [P''(h)]^2 \, dh = \boldsymbol{\gamma}^{\mathsf{T}}\mathbf{R}\boldsymbol{\gamma} = \left(\mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{g}\right)^{\mathsf{T}} \mathbf{R}\mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{g} = \mathbf{g}^{\mathsf{T}}\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{g} = \mathbf{g}^{\mathsf{T}}\mathbf{K}\mathbf{g}.$$

2.5.3 Semiparametrisen mallin estimaattori

Semiparametrisen mallin estimointikriteerin (15) neliösummaa voidaan haluttaessa painottaa painomatriisilla $\mathbf{W}$. Sen tulkinta ja muodostaminen on sama kuin painotetun pienimmän neliösumman menetelmässä eli sillä voitaisiin esimerkiksi poistaa tunnettu heteroskedastisuusrakenne. Sijoittamalla (28) kriteerin (15) rankaisutermin paikalle saadaan kahden ei-parametrisen muuttujan $j = 1, 2$ tapauksessa

$$\begin{aligned} \min S(\boldsymbol{\beta}, \mathbf{g}_j, \boldsymbol{\gamma}_j) = {} & (\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2)^{\mathsf{T}}\mathbf{W}(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2) \\ & + \alpha_1\boldsymbol{\gamma}^{\mathsf{T}}\mathbf{R}_1\boldsymbol{\gamma}_1 + \alpha_2\boldsymbol{\gamma}^{\mathsf{T}}\mathbf{R}_2\boldsymbol{\gamma}_2. \end{aligned} \tag{29}$$

Olemme kirjoittaneet estimointikriteerin siten, että kukin ei-parametrinen muuttuja saa oman järjestysmatriisinsa $\mathbf{N}_j$, vektorinsa $\mathbf{g}_j$ ja tasoitusparametrinsa α_j. Jatkossa myös muut aiemmin määritellyt matriisit ja vektorit merkitään kutakin eiparametrista muuttujaa koskevaksi vastaavalla tavalla. Minimointiongelmassa (29) on saatava ratkaisut parametreille β ja vektoreille g_j. Sijoittamalla (27) voidaan (29) ilmaista ilman derivaattoja,

$$\begin{aligned}\min S(\beta, g_1, g_2) &= (p - X\beta - N_1 g_1 - N_2 g_2)^T W (p - X\beta - N_1 g_1 - N_2 g_2) \\ &+ \alpha_1 (R_1^{-1} Q_1^T g_1)^T R_1 R_1^{-1} Q_1^T g_1 \\ &+ \alpha_2 (R_2^{-1} Q_2^T g_2)^T R_2 R_2^{-1} Q_2^T g_2. \end{aligned} \tag{30}$$

Lauseen (30) arvo minimoituu sen derivaatan nollakohdissa, sillä neliömuotojen kerroinmatriisit W ja $(R_j^{-1}Q^T)^T R_j R_j^{-1} Q^T$, $j = 1, 2$ ovat symmetrisiä ja positiivisesti definiittejä. Ensimmäiselle ei-parametriselle muuttujalle ratkaistaan lauseke,

$$\frac{\partial S(\beta, g_1, g_2)}{\partial g_1} = -2N_1^T W N_1 g_1 - 2N_1^T W (p - X\beta - N_2 g_2) + 2\alpha_1 Q_1 R_1^{-1} Q_1^T g_1 = 0. \tag{31}$$

Tämän ratkaisu g_1:n suhteen on

$$g_1 = \left(N_1^T W N_1 + \alpha_1 Q_1 R_1^{-1} Q_1^T\right)^{-1} N_1^T W (p - X\beta - N_2 g_2). \tag{32}$$

Pyrimme kuitenkin kirjoittamaan ratkaisun myös muuttujan toisen derivaatan suhteen, sillä laskuteknisesti on helpompaa ratkaista ensin g_1 γ_1:n funktiona ja sen jälkeen γ_1:n lauseke erikseen. Näin laskutoimitukset voidaan jakaa kahteen osaan ja splinin parametrien arvoon vaikuttava γ_1 saadaan laskennan oheistuotteena. Tästä tavasta käytetään nimeä *Reinsch-algoritmi.* Palataan sitä varten vielä rajoittamattomaan g_1:n ratkaisuun sijoittamalla (31):een rajoiteyhtälöt takaisin, eli $R_1^{-1} Q_1^T g_1 = \gamma$. Järjestelemällä termit uudelleen saadaan

$$N^T W N_1 g_1 = N^T W (p - X\beta - N_2 g_2) - \alpha_1 Q_1 \gamma_1. \tag{33}$$

Tarvittava lauseke g_1:lle saadaan suoraan edellisen ratkaisuna. γ_1:n lauseke saadaan ratkaistua sijoittamalla lausekkeesta (33) g_1 pois kertomalla puolittain käänteismatriisilla $(N_1^T W N_1)^{-1}$ ja Q_1^T:lla edestä ja sijoittamalla takaisin rajoiteyhtälöt saadaan

$$R_1 \gamma_1 = Q_1^T (N_1^T W N_1)^{-1} N_1^T W (p - X\beta - N_2 g_2) - \alpha_1 Q_1^T (N_1^T W N_1)^{-1} Q_1 \gamma_1.$$

g_1:n ja γ_1:n ratkaisut ovat Reinch-algoritmissa

$$g_1 = (N_1^T W N_1)^{-1} N_1^T (p - X\beta - N_2 g_2) - \alpha_1 (N_1^T W N_1)^{-1} Q_1 \gamma_1, \tag{34a}$$

$$\gamma_1 = (R_1 + \alpha_1 Q^T (N_1^T W N_1)^{-1} Q_1)^{-1} Q^T (N_1^T W N_1)^{-1} N_1^T W (p - X\beta - N_2 g_2). \tag{34b}$$

Vastaavasti voidaan johtaa lausekkeet $\mathbf{g}_2$:n ja γ_2:n suhteen. Laskeminen suoritetaan luonnollisesti sen omia matriiseja $\mathbf{Q}_2$, $\mathbf{R}_2$ ja $\mathbf{N}_2$, mutta yhteisiä painoja käyttäen.

$$\mathbf{g}_2 = (\mathbf{N}_2^T\mathbf{W}\mathbf{N}_2)^{-1}\mathbf{N}_2^T(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1) - \alpha_2(\mathbf{N}_2^T\mathbf{W}\mathbf{N}_2)^{-1}\mathbf{Q}_2\gamma_2 \tag{34c}$$

$$\gamma_2 = (\mathbf{R}_2 + \alpha_2\mathbf{Q}^T(\mathbf{N}_2^T\mathbf{W}\mathbf{N}_2)^{-1}\mathbf{Q}_2)^{-1}\mathbf{Q}^T(\mathbf{N}_2^T\mathbf{W}\mathbf{N}_2)^{-1}\mathbf{N}^T\mathbf{W}(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1). \tag{34d}$$

Myös γ_j:n ratkaisu on laskennallisesti raskas, mutta se voidaan laskea elementeittäin soveltamalla Choleskyn dekomponointia käännettävään matriisiin. Toisaalta nämäkin algoritmit ovat valmiiksi ohjelmoituna useissa numeerisen ja analyyttisen matematiikan ohjelmistoissa. Esimerkiksi Mathematican Solve-komento ratkaisee suhteellisen suuria harvoja yhtälöryhmiä.

Parametrivektorin $\boldsymbol{\beta}$:n ratkaisu saadaan käyttämällä jo edellä sovellettuja derivointisääntöjä,

$$\begin{aligned}
\frac{\partial S(\boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_2)}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}^T\mathbf{W}(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2) = 0 \\
&\Leftrightarrow \mathbf{X}^T\mathbf{W}\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^T\mathbf{W}(\mathbf{p} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2) = 0 \\
&\Leftrightarrow \boldsymbol{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(\mathbf{p} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2).
\end{aligned} \tag{34e}$$

Parametrivektori määräytyy PNS estimaattorina, jonka selitettävänä on ei-parametrisilta muuttujilta selittämättä jäävä osa hinnan vaihtelusta.

Semiparametrisen yhtälön "roughness penalty" estimaatit löydetään vektoreina $\boldsymbol{\beta}$, $\mathbf{g}_1$ ja $\mathbf{g}_2$, jotka simultaanisesti toteuttavat yhtälöt (34a) - (34e). Tässä yhteydessä käytämme ratkaisun löytämiseen ns. backfitting -menetelmää, vaikka numeerisesti vakaampiakin menetelmiä on olemassa, mutta backfitting käy mille tahansa muuttujamäärälle. Se konvergoi aina, mutta voi edetä alussa hyvin hitaasti.

Algoritmi etenee seuraavasti:

1. Rakennetaan järjestysmatriisi ja tarvittaessa painomatriisi.
2. Annetaan alkuarvot: Parametriselle osalle $\boldsymbol{\beta}^0 = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{p}$. Splinifunktiot asetetaan $\mathbf{g}_j^0 = 0$, merkitään $\boldsymbol{\eta}^0 = \left[\boldsymbol{\beta}^{0^T}, \mathbf{g}_j^{0^T}, \gamma_j^{0^T}\right]^T$
3. Ratkaistaan yhtälöt (34a) - (34e). Ratkaisuna saadaan $\boldsymbol{\eta}^1 = \left[\boldsymbol{\beta}^{1^T}, \mathbf{g}_j^{1^T}, \gamma_j^{1^T}\right]^T$.
4. Ellei konvergenssikriteeri täyty, saatu ratkaisu sijoitetaan alkuarvoksi ja ratkaistaan kohta 3.

Konvergenssikriteerinä käytämme parametrien summan suhteellista muutosta:

$$\Delta\eta = \left[\mid \eta^1 - \eta^0 \mid\right]^{\mathrm{T}} \mathbf{i} \;/\; \mid \eta^0 \mid^{\mathrm{T}} \mathbf{i}, \tag{35}$$

Kahden tai useamman tasoittajamuuttujan malli konvergoituu, mikäli tasoitettavat muuttujat eivät ole kollineaarisia tai kuten ei-parametrisessa ympäristössä ilmaistaan: käyrät ovat riippumattomia toisistaan. Muussa tapauksessa tulokset riippuvat lähtöarvoista. [HT90, s.123,140-1]

2.6 Tasoitusparametrin α valinta

Tasoitusparametri voidaan valita joko silmämääräisesti käyrän muodon perusteella tai jollakin harhaa ja ennustevarianssia minimoivalla lausekkeella. Jos valintakriteerinä käytetään tasoitusparametrin herkkyyttä aineiston vaihtelulle, on luonnollista minimoida ennustevirheen neliötä siten, että ennusteet lasketaan aineiston sisältä. Menetelmä tuntee nimen cross validation (CV). Periaatteena on ennustaa tulevaa selitettävän muuttujan p_i arvoa keinotekoisesti tiputtamalla havainto i pois aineistosta. Keskiarvo näin saavutetuista ennustevirheistä on CV-funktio

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \{p_i - \widehat{p}^{(-i)}(\alpha)\}^2, \tag{36}$$

missä $\widehat{p}^{(-i)}(\alpha)$ on ennustettu hinta havainnolle i. Tarkoituksena on sitten valita tasoitusparametri, joka minimoi CV:n. Käytännössä tämä edellyttää, että tasoitusparametrin arvoja joudutaan etsimään numeerisesti yli jonkin ennalta määritellyn alueen. Tämä on jo sellaisenaan melko työlästä ja lisäksi CV-funktio edellyttää n tasoitusta jokaiselle α:lle. Kuitenkin lineaarisen tasoitusmenetelmän ollessa kyseessä voidaan kirjoittaa:

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \left(\frac{p_i - \hat{p}_i}{1 - A_{ii}(\alpha)} \right)^2, \tag{37}$$

missä A_{ii} on diagonaalielementti matriisista $\mathbf{A}$ siten, että

$$\hat{\mathbf{p}} = \mathbf{A}\mathbf{p} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}_1\mathbf{g}_1 + \mathbf{N}_2\mathbf{g}_2.$$

$\mathbf{A}$ voidaan ratkaista kirjoittamalla kolme ensimmäisen asteen ratkaisua (niistä vain ensimmäinen (32) on edellä kirjoitettu auki) matriisiryhmänä,

$$\begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{N}_1 & \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{N}_2 \\ \mathbf{N}_1^{\mathrm{T}}\mathbf{W}\mathbf{X} & \mathbf{N}_1^{\mathrm{T}}\mathbf{W}\mathbf{N}_1 + \alpha_1\mathbf{K}_1 & \mathbf{N}_1^{\mathrm{T}}\mathbf{W}\mathbf{N}_2 \\ \mathbf{N}_2^{\mathrm{T}}\mathbf{W}\mathbf{X} & \mathbf{N}_2^{\mathrm{T}}\mathbf{W}\mathbf{N}_1 & \mathbf{N}_2^{\mathrm{T}}\mathbf{W}\mathbf{N}_2 + \alpha_2\mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}} \\ \mathbf{N}_1^{\mathrm{T}} \\ \mathbf{N}_2^{\mathrm{T}} \end{bmatrix} \mathbf{W}\mathbf{p}.$$

Ratkaisemalla parametrivektorin suhteen ja kertomalla vektorilla $\begin{bmatrix} \mathbf{X} & \mathbf{N}_1 & \mathbf{N}_2 \end{bmatrix}$ edestä saadaan matriisi $\mathbf{A}$ muotoon,

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{N}_1 & \mathbf{N}_2 \end{bmatrix} \times \begin{bmatrix} \mathbf{X}^T\mathbf{W}\mathbf{X} & \mathbf{X}^T\mathbf{W}\mathbf{N}_1 & \mathbf{X}^T\mathbf{W}\mathbf{N}_2 \\ \mathbf{N}_1^T\mathbf{W}\mathbf{X} & \mathbf{N}_1^T\mathbf{W}\mathbf{N}_1 + \alpha_1\mathbf{K}_1 & \mathbf{N}_1^T\mathbf{W}\mathbf{N}_2 \\ \mathbf{N}_2^T\mathbf{W}\mathbf{X} & \mathbf{N}_2^T\mathbf{W}\mathbf{N}_1 & \mathbf{N}_2^T\mathbf{W}\mathbf{N}_2 + \alpha_2\mathbf{K}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{N}_1^T \\ \mathbf{N}_2^T \end{bmatrix} \mathbf{W}.$$

$\mathbf{A}$ on $(n \times n)$ matriisi, eikä käytössämme ole oikotietä sen jäljen laskentaan, joten tyydymme approksimoimaan sitä parametristen muuttujien tasoittajan $\mathbf{P}$ ja ei-parametristen tasoittajien $\mathbf{S}_1$ ja $\mathbf{S}_2$ jälkien summalla. Tasoituksen $\hat{\mathbf{p}} = \mathbf{A}\mathbf{p}$ sijaan voimme eritellä kunkin muuttujaryhmän tasoituksen muodossa

$$\begin{aligned} \widehat{\mathbf{p}} &= \mathbf{N}_1\mathbf{g}_1 + \mathbf{N}_2\mathbf{g}_2 + \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{S}_1(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_2\mathbf{g}_2) + \mathbf{S}_1(\mathbf{p} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}_1\mathbf{g}_1) + \mathbf{P}(\mathbf{p} - \mathbf{N}_1\mathbf{g}_1 - \mathbf{N}_2\mathbf{g}_2). \end{aligned}$$

$\mathbf{S}_1$ voidaan ratkaista suoraan (32) :sta kertomalla edestä $\mathbf{N}_1$:llä,

$$\mathbf{S}_1 = \mathbf{N}_1\left(\mathbf{N}_1^T\mathbf{W}\mathbf{N}_1 + \alpha_1\mathbf{Q}_1\mathbf{R}_1^{-1}\mathbf{Q}_1^T\right)^{-1}\mathbf{N}_1^T\mathbf{W}.$$

Vastaavasti voidaan ratkaista $\mathbf{S}_2$. Parametrisen osan tasoittaja saadaan ratkaisuna (34e):sta,

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}.$$

Hastien ja Tibshiranin simulointitulokset [HT90, sivu 129] viittaavat, että näiden matriisien jälkien summa approksimoi hyvin matriisin $\mathbf{A}$ jälkeä, mikäli selittävät eivät ole vahvasti korreloituneet tai tasoitusparametrit ovat erittäin pieniä.

Additiivisessa mallissa tasoitusparametrit on valittava yhtäaikaa, joten optimaalisen parametriyhdistelmän valinta vaatisi kaksiulotteisen numerokentän haarukointia. Laskentatoimenpiteet ovat monimutkaisia eivätkä automaattisen valinnan tulokset ole aina luotettavia varsinkin, kun joudumme approksimoimaan $\mathbf{A}$ matriisin jälkeä. (ks. tarkemmin keskustelua [HT90, sivut 42-3, 47-9, 52]).

2.7 Tilastollinen päättely

Tilastollinen päättely tuloksista on vielä tuntemamme kirjallisuuden perusteella vähän tutkittu alue. Green ja Silverman [GS94] välttävät keskustelemasta koko aiheesta vedoten sen tutkimuksen keskeneräisyyteen. Hastie ja Tibshirani [HT90] kuitenkin esittävät testausmenettelyjä, joiden on osoitettu ainakin joissakin simuloinneissa tuottavan luotettavia tuloksia. Esitämme tässä yhden tavan tehdä ei-parametristen mallien tilastollista päättelyä. Esitettyihin arvioihin testien koosta tai voimakkuudesta tulee kuitenkin suhtautua varauksella.

Tähänastinen analyysimme on perustunut puhtaasti reaalilukujen matematiikkaan. Testaus edellyttää tilastollista mallia, joka tässä yhteydessä on yksinkertaisin mahdollinen. Oletetaan, että semiparametrisella mallilla on satunnainen virhetermi ε, jonka odotusarvo on nolla vakiovarianssilla, $var(\varepsilon) = \sigma^2$.

Ei-parametrisessa estimoinnissa mallin paremmuuden arviointi perustuu poikkeaman ja vapausasteiden analysoitiin. Poikkeaman tehtävä ei-parametrisessa estimoinnissa on sama kuin residuaalien neliösumman parametrisessa estimoinnissa:

sovitteen hyvyyden arviointi ja mallien vertailu. Normaalijakaumaoletuksella se itse asiassa pelkistyy residuaalien neliösummaksi. Poikkeaman ohella tarvitaan tietoa vapausasteista, jotka kertovat tasoituksen suuruudesta ja sen aiheuttamasta uhrauksesta mallille. Vapausaste $df = 0$ tarkoittaa, että muuttujaa ei ole. Lineaarisille termeille $df = 1$, ja mitä suurempi luku on sitä tarkempi sovite eli epätasaisempi käyrä. [HT90, sivut 52, 155-6, 259-60] Tasoituksen vapausasteet lasketaan klassisen lineaarisen mallin tavoin projisio- tai tässä tapauksessa tasoitusmatriisin **A** jäljestä. Tässäkin tapauksessa approksimoimme tätä muuttujaryhmien tasoittajien jälkien summalla. Nekin ovat $(n \times n)$ matriiseja, mutta niiden nauhamainen rakenne sallii taloudellisen tavan käsitellä niitä. Algoritmiä on kuvattu Greenin ja Silvermanin [GS94, ss. 33-35, ja *errata*] teoksessa.

Tasoituksen määrää kuvaavaa vapausastetta merkittävämpi mittari mallin ominaisuuksien vertailuun on virhetermin vapausasteet $df^{err} = n - tr(2\mathbf{A} - \mathbf{AA})$. Lineaarinen malli on aina erikoistapaus ei-parametrisesta, $(\alpha \to \infty)$. Yksittäisen muuttujan tai rajoituksen vaikutus tasoitukseen näkyy virhetermin vapausasteiden muutoksena

$$df_j^{err} = tr(2\mathbf{A} - \mathbf{AA}) - tr(2\mathbf{A}_j - \mathbf{A}_j\mathbf{A}_j) \tag{38}$$

missä $\mathbf{A}_j$ on tasoitusmatriisi muuttujan j osalta rajoitetusta regressiosta.[13] Mallien vertailuun voidaan käyttää vapausasteilla korjattua residuaalien neliösumman suhteellista muutosta. Testisuureella

$$F = \frac{(RSS_j - RSS)/df_j^{err}}{RSS/df^{err}}, \tag{39}$$

on ainakin asymptoottisesti F jakauma vapausasteilla (df_j^{err}, df^{err}). RSS_j on rajoitetun mallin residuaalien neliösumma, n havaintojen lukumäärä. Yhtälössä (38) esiintyvien matriisin jälkien laskeminen on liian suuri urakka kotitietokoneille. Sitä yksinkertaistaan Green ja Silvermanin [GS94] ehdottamalla approksimaatiolla $df_{err} = n - tr(\mathbf{A})$ tai Hastie ja Tibishiranin [HT90] ehdotuksella $1,25(n - df_{err}) - 0,5$. Ilman momenttikorjauksia testin 95 prosentin koko on todellisuudessa 90 ja 93 prosentin välillä ([HT90, sivut 66,158]). Testimenettelyssä joudutaan approximoimaan testin vapausasteita kahteen kertaan ja lisäksi varsinainen testisuurekin vain approksimoi F-jakaumaa. On selvää, ettei tuloksille voi antaa suurta painoa.

Mallin parametrisen osan tilastolliseen testaukseen käytämme kovarianssimatriisia

$$cov[\hat{\beta}] = (\mathbf{X}^T\mathbf{X})^{-1} s^2, \tag{40}$$

missä

$$s^2 = \sum_{i=1}^{n} (p_i - \hat{p}_i)^2 / df_{err}. \tag{41}$$

[13] Toisin kuin lineaarisessa rajoitteessa, ei-parametrinen osa kuluttaa vaihtelevan määrän vapausasteita.

3 Joensuun asuntomarkkinat

Tässä kappaleessa sovellamme edellisessä kappaleessa rakennettua estimointikehikkoa hedoniselle hintafunktiolle. Havainnollistamme muutamia estimaattorin laskentaan liittyviä piirteitä, kuten cross validation ja rankaisutermin laskentaa. Nämä tietenkin kiinnostavat pääasiassa vain ei-parametrisiin tekniikoihin perehtyneitä lukijoita. Kuutiosplinin ja semiparametrisen mallin soveltamisesta kiinnostuneille tarjoamme mahdollisuuden käyttää kirjoittamaamme Mathematica koodia, joka on noudettavissa vapaasti sivuilta http://www.vatt.fi tai pyydettäessä suoraan allekirjoittaneilta. Valittu ohjelmisto ei ole optimaalinen alusta numeeriselle laskennalle, mutta toisaalta helpompi käyttää kuin monet erityisohjelmat. Laskentatehoa voi helposti kasvattaa esimerkiksi tieteellisen laskentapalvelun koneilla olevalla SPlus ohjelmistolla, joka sisältää valmiita rutiineja ei-parametriseen estimointiin.

3.1 Aineisto

3.1.1 Yleinen kuvaus

Sekä Sari Hukkanen [Huk90], että Antti Miettinen [Mie95] ovat opinnäytetöissään Joensuun yliopiston taloustieteen laitoksella keränneet aineistoa Joensuun asuntomarkkinoilta. Hukkasen -pro gradu työssä kerättiin esimmäistä kertaa kattavampi otos, joka perustuu Suomen kiinteistövälittäjäin liiton keräämään kauppakohtaiseen aineistoon 1985-1989. Sitä laajennettiin edelleen Miettisen ja Kyllösen pro graduissa [Kyll98] aina vuoteen 1997 (Helmikuu) asti. Aineistosta puuttuu havaintoja vuodenvaihteen 1994 molemmin puolin. Alkuperäisessä aineistossa ovat mukana kerros-, rivi- ja omakotitalot. Omakotitalot on jätetty tämän työn ulkopuolelle, koska ne oletettavasti sisältävät paljon aineiston ulkopuolisia talotyyppiin liittyviä ominaisuuksia, kuten takkojen, uunien ja autotallin olemassa olo sekä tontilla olevien puiden lukumäärä. Rivitaloasunnot on otettu mukaan, jotta havaintojoukko olisi alueellisesti mahdollisimman monipuolinen, ja ympäristömuuttujiin saataisiin olennaista vaihtelua. Esimerkiksi vesistön äärellä ja kaukana keskustasta sijaitsevia kerrostaloja on rajoitetusti, kun sitä vastoin rivitaloja on lukuisia. Toisaalta ne sisältävät sellaisia erityisominaisuuksia, kuten pieni pihamaa. Näitä kaikkia erityisomaisuuksia mittaa vain rivitalo-dummy.

Aineistosta puuttuu yhden suuren välittäjän välittämät kaupat ja kaikki ilman välittäjää tehdyt kaupat. Kattavuus vapaarahoitteisesta kaupasta on noin 40 prosenttia. Sosiaalisesti rahoitettu asuntotuotanto muodostaa näiden lisäksi oman erityisen markkinansa.

Vuosien 1994-97 aineistossa puuttuvien tietojen ongelmaa on ratkaistu kolmivaiheisella proseduurilla. Ensimmäiseksi on katsottu, missä osoitteessa puuttuva tieto on ja korvattu se saman osoitteen olemassa olevalla havainnolla. Näin on usein onnistuneesti menetelty rakennusvuoden kohdalla. Toiseen vaiheeseen turvaudutaan silloin, kun samaa osoitetta eli samaa taloa ei esiinny muissa havainnoissa tai saman osoitteen arvoilla on vaihtelua. Tällöin puuttuvat tiedot korvataan olemassa olevien havaintojen keskiarvolla. Tällä tavoin on täytetty muutamien vastike- ja sijaintikerros tietojen puuttuvat havainnot. Se ei tosin paranna kyseisten muuttujien osalta estimointituloksia, ja tuloksena on vain huono sovite [Gre93, sivu 276].

Viimeisenä mahdollisuutena yksinkertaisesti poistetaan havainnot, joissa on puuttuvia tietoja. Näin toimittiin mikäli puuttuva tieto oli selitettävässä muuttujassa ja usealle puuttuvalle rakennuksen ikä ja asunnon kunto tiedolle. Karsimisen seurauksena havaintojen lukumäärä laskee vuosien 1993 - 97 aineistossa 1010:een. Kaikkien operaatioiden oletuksena on, että puuttuvat tiedot eivät riipu systemaattisesti selitettävästä ilmiöstä. Miettisen aineistossa kaikki puuttuvia tietoja sisältäneet kaupat oli jo jätetty pois

3.1.2 Selitettävä muuttuja

Selitettävänä muuttujana on asuntojen kulloistakin markkina-arvoa kuvaava nimellinen asunnon velallinen myyntihinta. Sitä ei siis muuteta nykyarvoon hintaindeksien avulla. Hintaindeksit ovat aggregoituja ja estimoituja suureita, joiden käyttäminen markkina-arvon muuttamisessa reaaliarvoiseksi aiheuttaisi informaation menettämistä. Tällöin jäisi käyttämättä mahdollisuus omalla mallilla ja paikalliset olot kattavalla aineistolla selvittää asuntomarkkinoiden kulloisenkin tilan vaikutus hintaan.

Osalla asunnoista on yhtiövelkaa, osalla taas ei. Periodilta 1985-93 ei ollut mahdollista kattavasti selvittää velan määrää tai edes sen olemassaoloa. Toisaalta käyttämämme yhtiövastike-muuttuja selittävänä tekijänä sisältää myös yhtiölainojen kustannukset.

3.1.3 Selittävät muuttujat

Miettisen pro gradussa [Mie95] estimoitiin samankaltainen hedoninen malli, mutta täysin parametrisena. Siinä käytiin läpi melko kattavasti erilaisia asuntojen ominaisuuksiin ja asuinalueisiin liityviä selittäviä muutujia, joten valitsemme selittävät muuttujat näihin tuloksiin nojaten.

Asuntojen nimellishintojen vaihtelun selittävinä muuttujina käytetään asunnon rakenteellisia ja sijainnillisia tekijöitä. Aikatrendi on mukana huomioimassa asumisen inflaation. Miettisellä merkitseviä rakenteellisia muuttujia olivat asunnon pinta-ala, sauna-dummy, rakennuksen ikä ja yhtiövastikkeen suuruus. Ne ovat mukana myös tässä työssä lukuunottamatta saunaa periodille 1994 - 97. Lisäksi mukaan on otettu merkitsemättömän kertoimen saanut muuttuja asunnon sijaintikerros sekä uudemmassa aineistossa asunnon kunto. Hedonisen hintafunktion muuttujavalintaa ja funktiomuotoa ajatellen kaikki yllämainitut muuttujat asunnon ikää ja sijaintikerrosta lukuun ottamatta ovat varsin ongelmattomia. Niiden vaikutus hintaan voidaan olettaa monotonisen kasvavaksi

Asunnon iän vaikutusta asunnon arvoon on hankala määritellä etukäteen. Miettisen tutkimuksessa iän vaikutus oli negatiivinen: Hinnan ikäjoustoksi saatiin -0.1. Ikävaikutus ei kuitenkaan ole välttämättä koko ajan samansuuntainen. Syyt tällaiseen ilmiöön voivat olla kuluttajien preferensseihin liittyviä. Esimerkiksi 1960 -luvulla rakennettua asuinkerrostaloa voidaan pitää miellyttävämpänä kuin 1970 -luvulla valmistunutta. Toinen ilmeinen ikään liittyvä tekijä on asuntojen peruskorjaus. Vanha peruskorjattu talo voi olla arvokkaampi kuin uudempi, mutta heikkokuntoinen. Jotta edellä mainitut ilmiöt näkyisivät parametrisessa mallissa merkitsevinä, niiden täytyy vaikuttaa säännöllisesti, eli asuntoja peruskorjataan pääpiirteissään samoissa i'ssä ja arkkitehtoniset ratkaisut tietyllä ajanjaksolla ovat samantyyppisiä. Emme esitä mitään hypoteesia asunnon iän vaikutuksen suunnasta, vaan liitämme sen malliin ei-parametrisena muuttujana.

Toisena ei-parametrisena muutujana käytämme kaupantekovuotta. Tarkoituksena on tuottaa mahdollisimman hyvä "deflaattori" asuntojen hinnalle.

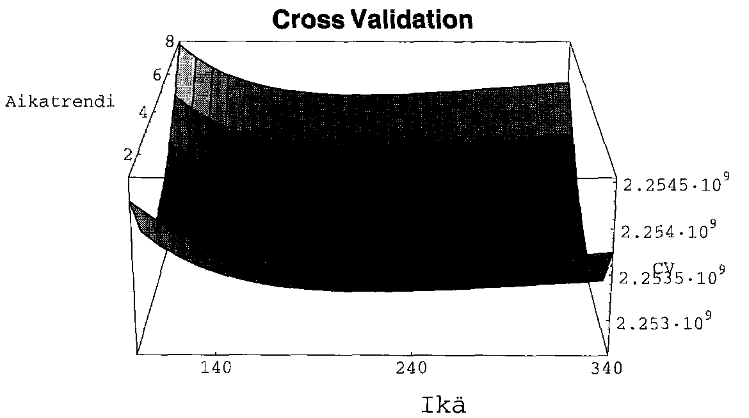
Sijaintikerroksen vaikutuksesta hintaan ei ole Miettisen tutkimuksessa näyttöä, joten se soveltuisi myös hyvin ei-parametrisessa muodossa käsiteltäväksi muuttujaksi. On mahdollista, että korkealla asumista arvostetaan näköalojen ja puhtaamman ilman takia. Toisaalta siihen liittyy käytännöllisiä, muuttamiseen ja liikkumiseen liittyviä ongelmia etenkin, jos talossa ei ole hissiä. Ylivoimaisesti suurin osa Joensuun kerrostaloista on kolmesta neljään kerroksisia, joten kerroksista voitaisiin kohtuudella laatia myös dymmy-muuttuja. Sijaintikerros jätetään kuitenkin sellaisenaan parametriseksi muuttujaksi.

Miettisen tutkimuksen merkitsevistä sijainnillisista muuttujista on tähän työhön otettu mukaan yleisiä teitä pitkin mitattu keskustaetäisyys ja ympäröivien palvelujen vaikutusta hintaan kuvaava myymälärakennusten kerrosalojen osuus kaikista kerrosaloista sekä merkitsemättömän kertoimen saaneet muuttujat: vesistön ja rautatien tai vilkkaasti liikennöidyn tien läheisyys. Liikenteen vaikutusta kuvataan yhdellä dummy-muuttujalla. Vesistön vaikutusta mitataan kahdella dummy-muuttujalla. Veden läheisyyttä kuvaava dummy saa arvon yksi, kun asunto on alle 300 metrin päässä rannasta, riippumatta vesinäköalasta. Vesinäköalaa kuvaava dummy saa arvon yksi, jos asunto on alle 300 metrin päässä rannasta ja rakennuksen ja rannan välissä ei ole muita rakennuksia. Sekä ranta että liikenne dummyssä huoneisto voi sijaita rakennuksen rauhallisemmalla puolella tai poispäin vesistöstä. Oletamme, että näiden tekijöiden vaikutus kohdistuu suhteellisen tasaisena koko rakennukseen.

Jo asunnon iän kohdalla jouduttiin miettimään havaitun muuttujan taustalla vaikuttavia tekijöitä. Kaikki kiinnostavat asunnon ominaisuudet eivät ole suoraan havaittavissa, vaan joudumme käyttämään näiden taustamuutujien mittaamiseen proxy-muutujaa. Esimerkiksi asunnon "suuruutta" voidaan mitata huoneiden lukumäärällä tai asunnon pinta-alalla. Molemmat mittaavat hieman eri näkökulmaa "asunnon suuruuteen" ja niiden pitäisi olla edustettuina selittävissä muuttujissa. Usean, osin samaa tekijää mittaavan muuttujan käyttö regressioyhtälössä ei ole kovin hedelmällistä, sillä voimakkaasti korreloivien muuttujien vaikutusta ei pystytä erittelemään. Tätä multikollineaarisuusongelmaa lievennetään jättämällä estimoitavan yhtälön parametriseen osaan yksin asunnon pinta-ala mittaamaan asunnon kokoa. Valinta perustuu koe-estimointeihin ja muitakin vastaavia muuttujapareja on jouduttu testaamaan mallin estimointivaiheessa.

3.2 Hedoninen hintafunktio

Käytössämme oleva muuttujavalikoima ei ole koko havaintoajanjaksolta yhtenäinen. Vuosilta 1985 - 93 emme tiedä asunnon kuntoa ja vuosilta 1994 - 97 emme tiedä onko asunnossa saunaa. Käytettyjen selittävien muuttujien lisäksi periodien vertailukelpoisuuteen vaikuttaa rankaisutermin arvo. Yhtenäisyyden vuoksi estimoimme aluksi optimaaliset rankaisutermit yhdistetystä aineistosta ja samalla tarkastelemme mallispesifikaatiota suhteessa rajoitettuun, lineaariseen malliin sekä tutkimme virhetermin käyttäytymistä. Jäljempänä asuntomarkkinoiden analyysit perustuvat



Kuvio 4: Cross Validation mittarin simuloinnit

periodien omiin funktioihin kuitenkin siten, että rankaisutermien estimaatit tulevat yhdistetystä aineistosta.

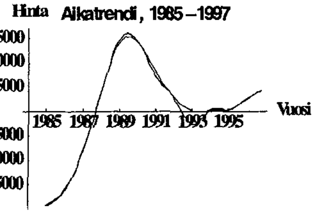
3.2.1 Estimoinnit yhdistetyllä aineistolla

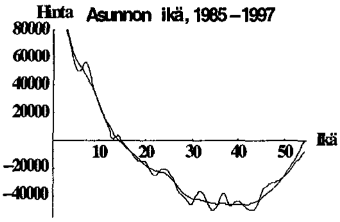
Semiparametrisen mallin tulokset riippuvat hyvin monesta tekijästä, mutta tässä yhteydessä olemme kiinnostuneet erityisesti rankaisutermin vaikutuksesta. Valintaa voidaan ohjata edellä esitettyjen Cross Validation mittarin ja vapausasteiden määrän perusteella. CV-mittari muuttuu aina samaan suuntaan ennustevirheen neliön odotusarvon kanssa. Myös mallin vapausasteiden määrä muuttuu tietyin edellytyksin samaan suuntaan kuin CV-mittari. Nämä mittarit eivät ole kuitenkaan ongelmattomia, sillä CV saa helposti lokaaleja minimejä ja varsinkin additiivisessa mallissa minimoivat rankaisutermit on simuloitava moniulotteisesta valintajoukosta. Olemme simuloineet kahden ei-parametrisen muuttujan semiparametrista mallia

$$\begin{aligned} P = vakio * \beta_1 + ala * \beta_2 + kerros * \beta_3 + vastike * \beta_4 + kesk.et * \beta_5 \\ + tie * \beta_6 + ranta * \beta_7 + g\,(aikatrendi) + g\,(ikä) + \epsilon \end{aligned} \tag{42}$$

koko aineistolla muutellen rankaisutermien arvoja. Mallissa (42) ϵ on vapaa virhetermi, jonka ominaisuuksia käsittelemme myöhemmin. Kuviossa *4* on piirretty mittarin minimi $\alpha_{aika}, \alpha_{ikä} = [4, 220]$. Käytämme sitä jatkossa kaikissa raportoiduissa tuloksissa

Rankaisutermin vaikutusta voidaan parhaiten havainnollistaa graafisesti. Kuviossa *5* on piirretty estimoidut splinit rankaisutermien yhdistelmille $(\alpha_{aika}, \alpha_{ikä}) = (0.5, 5)$ ja $(\alpha_{aika}, \alpha_{ikä}) = (4, 220)$. Suhdanteiden vaikutus läpäisee Joensuussa koko asuntomarkkinat, joten suhteellisen pieni rankaisutermi verrattuna iän rankaisuun tuottaa melko häiriöttömän käyrän. Rankaisutermi ei vaikuta oleellisesti käyrän keskeisiin ominaisuuksiin—ainoa olennainen asia, joka jää huomaamatta voimakkaammassa rankaisussa on vuoden 1993 pohjanoteeraus hinnoissa. Asunnon ikähavain-





Kuvio 5: Ajan ja asunnon iän vaikutus asunnon hintaan ranskalaisittain. Polveikkaamman käyrän rankaisutermit ovat $(\alpha_{aika}, \alpha_{ikä}) = (0.5, 5)$*, tasaisempi käyrä on optimaalisesti tasoitettu* $(\alpha_{aika}, \alpha_{ikä}) = (4, 220)$.

noilla on huomattavasti epävarmempi tai vaihteleva vaikutus hintaan. Tämä näkyy voimakkaasti polveilevana käyränä pienellä tasoituksella.

Semiparametrisen spesifikaation selitysvoimaa voidaan testata esittämällä nollahypoteesi lineaarisen mallin uskottavuudesta. Lineaarinen malli voidaan tulkita rajoitetuksi versioksi semiparametrisesta, joten hypoteesia voidaan testata testisuureella (39). Tulokset on koottu taulukkoon *1*. Niiden perusteella voidaan hylätä

Taulukko 1: Yhteenveto mallien estimointi- ja testaustuloksista

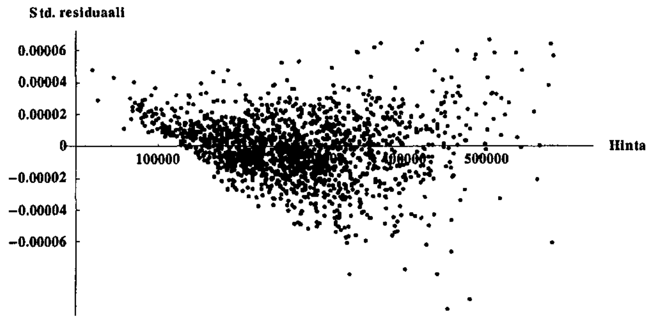
malli	lineaarinen	Semiparametrinen
iteraatiot	–	14
vapausasteet	9	28, 17
RSS	6.43×10^{12}	3.38×10^{12}
F-testisuure	–	$F(20, 1684) = 75, 12$

nollahypoteesi rajoitetun (lineaarisen) mallin uskottavuudesta suhteessa semiparametriseen. Aineisto ainakin yhdistettynä ja sen tarkemmin selittäviä muuttujia erittelemättä tukee semiparametrista spesifikaatiota.

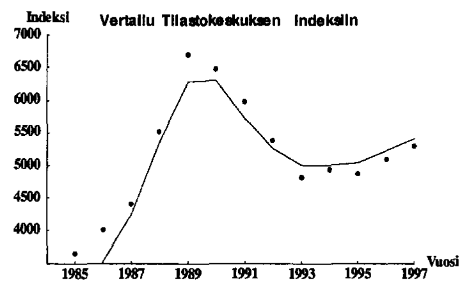
Virhetermiä koskevien oletusten, $E(\epsilon_i) = 0$ ja $Var(\epsilon_i) = vakio$, voimassaoloa tutkitaan alla olevalla residuaalikuviolla. Se on tehty piirtämällä kuhunkin havaintoon liittyvän sovitteen $\mathbf{p} = \mathbf{X}\beta + \mathbf{N}_1\mathbf{g}_1 + \mathbf{N}_2\mathbf{g}_2$ arvot x-akselille ja standardoidut residuaalit y-akselille. Residuaalit on standardoitu kaavalla $st.e_i = \frac{e_i}{\sqrt{(s^2(1-A_{ii}))}}$, jossa A_{ii} on yhteenlaskettujen tasoitusmatriisien diagonaalielementti. [Gre93, p. 288].

Kuvion perusteella jonkinasteista heteroskedastisuutta lienee olemassa, koska residuaalien varianssi kasvaa systemaattisesti hintahaitarin ääripäissä. Painomatriisin käyttäminen estimointikriteerissä voisi hyvinkin poistaa heteroskedastisuuden, mutta samalla menetettäisiin tärkeää tietoa mitattavasta ilmiöstä. Jätämme residuaalit käsittelemättä ja tyydymme harhaisiin varianssiestimaatteihin.

Selitettävä muuttuja jätettiin estimoineissa deflatoimatta, sillä ajan vaikutus haluttiin estimoida suoraan aineistosta. Kuviossa *7* verrataan saatuja tuloksia tilastokeskuksen laatimaan Joensuun vapaarahoitteisten asuntojen neliöhinnan indek-



Kuvio 6: Sovitteen ja standardoitujen residuaalien riippuvuus.



Kuvio 7: Vertailu tilastokeskuksen indeksin ja splinin välillä. Viiva kuvaa splinistä laskettua indeksiä ja pisteet ovat tilastokeskuksen raportoimia lukuja.

siin [Til]. Vertailu ei ole täysin yksiselitteistä. Estimoitu splini-käyrä kertoo asunnon markkinahinnan muutoksen ajanhetkellä absoluuttisena markkamääränä. Olettamalla asunnon koon pysyneen samana tarkasteluajalla, käyrä on muutettu vertailukelpoiseksi indeksiksi suoraan suhteuttamalla muutokset asuntojen keskihintaan. Splini-käyrä näyttäisi kuvaavan melko tarkoin samaa asiaa kuin tilastokeskuksenkin laskema indeksi.

3.2.2 Estimoinnit periodeille 1985-93 ja 1994-97

Estimoinneissa rivitalo-dymmy jää molemmissa aineistoissa selvästi merkityksettömäksi samoin kuin palveluiden saatavuus alueella. Miettisen [Mie95] tutkimuksessa palvelut saivat merkitsevän kertoimen, mutta käytetty muuttuja ei ole kaikilta osin sama. Myöskään hissillä ei ole merkitsevää kerrointa uudemmassa aineistossa. Tämä johtunee sijaintikerrosmuuttujan aiheuttamasta yhteisvaihtelusta. Vesistön läheisyyttä kuvaava dummy on liian karkea, sillä se saa negatiivisen kertoimen ja vieläpä merkitsevänä, jos kaikki muuttujat otetaan malliin. Estimoitavaan yhtälöön jätetään vain vesinäköalaa kuvaava ranta-dummy. Yhteenveto tuloksista on esitetty taulukossa *2* ja parametriestimaatit taulukossa *3*. Rakennuksen iän ja ajan vaikutus on eritelty kuvioissa *9* ja *8*.

Taulukko 2: Estimointien tunnuslukuja periodeilta

	Vuodet 1985-93	Vuodet 1994-97
Rankaisutermi	$(\alpha_{aika}, \alpha_{ikä}) = (4, 220)$	$(\alpha_{aika}, \alpha_{ikä}) = (4, 220)$
Vapausasteet	23.12	22.13
Residuaalien neliösumma	9.39×10^{11}	2.210×10^{12}
Iteraatiot	15	10

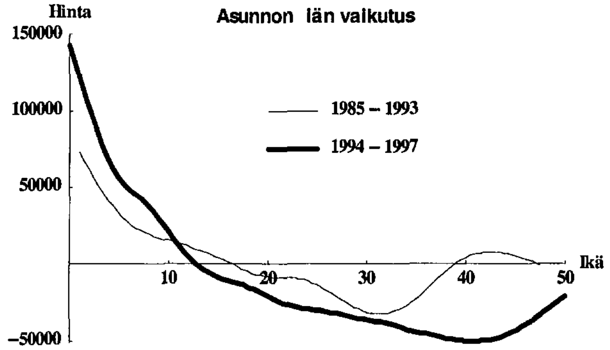
Taulukko 3: Estimoidut parametrit periodeilta

	Vuodet 1985-93		Vuodet 1994-97	
muuttuja	kerroin	t-testi	kerroin	t-testi
VAKIO	124102	14.38	176669	19.27
ALA	3526.6	47.82	3317.48	47.51
KERROS	416.39	0.29	8518.32	4.9
KUNTO			-24678.1	-7.43
SAUNA	44912.5	10.81		
VASTIKE	-2544.47	-2.98	-2448.23	-4.85
KESK.ET	-29.7	-17.87	-28.4	-29.20
LIIKENNE	-2644.15	-0.86	2447.07	0.60
RANTA	30852.4	4.21	33641.9	3.64

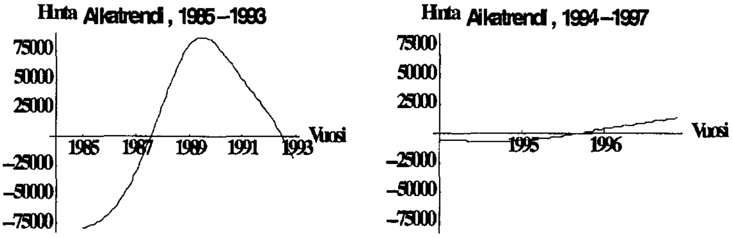
Estimoitujen parametrien stabiilisuutta vertaillaan yleensä Chow-testillä. Sen sovellettavuutta semiparametriseen malliin ei ole dokumentoitu, joten jätämme sen tarkemman pohdiskelun tulevaisuuteen. Osittaisen kuvan parametrien ajallisesta stabiilisuudesta saa tarkastelemalla parametrien luottamusvälejä. Lineaarisen osan parametreistä kaikki muut paitsi vakion ja sijaintikerroksen parametrien 95 %:n luottamusvälit ovat päällekkäiset, joten muutoksia asuntojen ominaisuuksien vaikutuksessa ei pystytä tilastollisesti havaitsemaan. Kerroksen lisääntyminen yhdellä nostaa asunnon hintaa karkeasti 8500 markkaa laman jälkeen. Kun kerrosparametri ei poikennut ennen lamaa tilastollisesti nollasta, sen vahva vaikutus jälkimmäisellä periodilla voi hyvinkin kuvastaa asuntomarkkinoiden rauhoittumista. Emme voi kuitenkaan sulkea pois tilastointitarkkuudessa tapahtuneita muutoksia. Vanhemman aineiston perusteella sauna nostaa asunnon hintaa karkeasti 45 000 markkaa. Uudemmassa aineistossa asunnon kuntoluokan lasku laskee hintaa 25 000 markkaa.

Vakiotermin tulkinta on hieman ongelmallinen, sillä se indentifioituu vain yhdessä ei-parametristen osien kanssa. Käyriin voidaan tehdä tasomuutoksia ja kompensoida ne muutoksella vakiotermissä. Siten ei-parametrista osaa voidaan (näissä estimoinneissa) tulkita vain selittävän muuttujan vaikutuksen muutoksina, eli marginaalivaikutuksina, ja vakiota ehdollisena splini-käyrien estimoidulle korkeudelle (tasolle).

Asunnon iän hintavaikutus (Kuva *8*) on hyvin voimakas ensimmäiset 10-15 vuotta. Periodilta 1994 - 97 estimoituna uuden asunnon hinta laskee ensimmäisen vuosi-



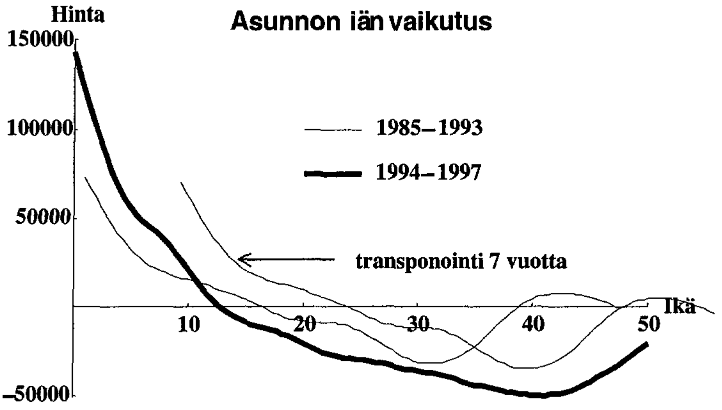
Kuvio 8: *Asunnon iän vaikutus myyntihintaan.*



Kuvio 9: *Ajan vaikutus asunnon hintaan*

kymmenen aikana 120 000 markkaa ja sitä seuraavan viiden vuoden aikanakin vielä karkeasti 50 000 markkaa lisää. Asuntojen laadullisella muutoksella tätä hintamuutosta ei voi selittää kokonaan, sillä asunnon kunto, vastike, sijainti ja pinta-ala on jo huomioitu muissa regressoreissa. Hintamuutoksen vaikuttajaksi jää siten rakenteiden kulumisen lisäksi muut ikään ja asunnon varusteluun liittyvät tekijät. Jälkimmäisellä periodilla uudemmissa asunnoissa yleistynyt sauna voi olla tämänlainen merkittävä tekijä, mutta ei pysty yksin selittämään koko arvon alennusta. Uusasunnon ostaja kuluttaa asumisellaan ensimmäisen 15 vuoden aikana asuntonsa velallisesta hinnasta karkeasti 100 000 markkaa. Tämä summa ei palaudu hänelle asuntoa myydessä tai vaihdettaessa, vaan hän on valmis maksamaan tämän preemion asunnon uutuudesta. Lisäksi tulevat mahdolliset yhtiövelan maksut. Jos uustuotannon hinta nousee muun kuin asumisen laadun kohoamisen seurauksena, esimerkiksi rakennuskustannusten tai tonttimaan kallistumisen seurauksena, ei hintojen nousu hedonisessa tarkastelussa siirry vanhojen asuntojen markkinoille. Eksogeeniset, vain osaan markkinoita kohdistuvat shokit jäävät vaikuttamaan vain näihin hintoihin. Jos ostajat eivät näitä hintoja hyväksy, rakentajien täytyy laskea saamaansa katetta muista rakennetuista ominaisuuksista.

Toinen mielenkiintoinen ilmiö asunnon iän vaikutuksessa on eri periodeilta esti-

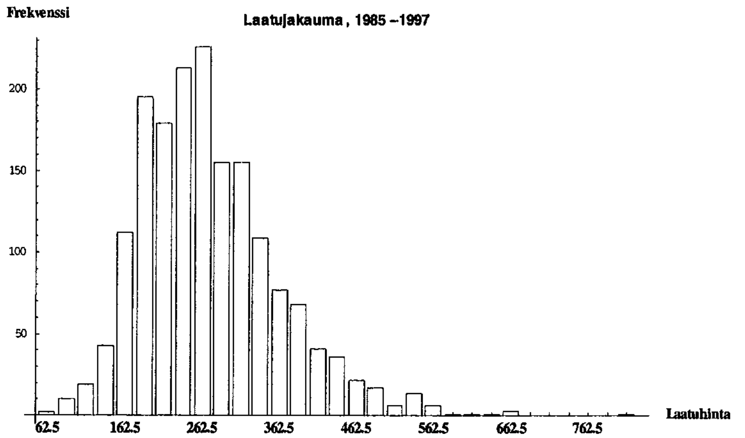


Kuvio 10: *Vuosien 1985-1993 estimaatin transponointi. siirtymä korjaa erot asuntojen iässä eri estimointiperiodeilla.*

moitujen käyrien samankaltaisuus. Piirretyt käyrät ovat kaltevuudeltaan liki yhteneviä, jos otamme huomioon erot havaintoperiodeissa, eli siirrämme periodin 1985 - 93 käyrää horisontaalisesti 5 - 7 vuotta oikealle, kuten kuviossa *10* on tehty. Tämä vastaa karkeasti rakennuskannassa tapahtunutta muutosta estimointiperiodien välillä. Käyrien erot estimointiperiodeilla ja niiden yhtyminen periodien eroa korjattaessa viittaa siihen, että asunnon ikä sellaisenaan ei ole ratkaiseva tekijä asunnon hinnan määrityksessä, vaan myös rakentamisajankohta merkitsee. 70-luvun lopulla ja sitä ennen rakennetut talot—1994 - 97 käyrältä yli 15-vuotiaat rakennukset—menettävät hintaansa vähemmän vuotta kohden kuin nuoremmat rakennukset. Ilmeisesti ostajien mieltymykset vanhasta ja uudesta muuttuvat juuri vuosikymmenen taitteen rakennusten kohdalla. Koettu lisähaitta ikääntymisestä ei ole enää niin tärkeä tekijä, jos asunto on 70-luvulla tai sitä ennen rakennettu. Vastaavasti 80- ja 90-luvulla rakennettua asuntoa pidetään uutena tai modernina ja sen ikääntyminen painaa paljon sen markkina-arvossa.

Tutkimuksen alkuvaiheessa kuvittelimme pystyvämme paikantamaan rakennusten peruskorjaukseen liittyvän syklin hinnassa. Tuloksista ei voi tätä päätellä. Ensimmäisellä periodilla estimoituna hinta nousee vasta yli 30 vuotta vanhoille asunnoille ja jälkimmäisellä periodilla yli 40 vuotiaille[14]. Jälleen hintamuutos näyttäisi olevan sitoutunut rakennusvuoteen, ei asunnon ikään. Saman vuosikerran asunnot eivät voi olla jatkuvasti peruskorjauksen tarpeessa, joten vanhojen, 50-luvulla rakennettujen kerros- ja rivitaloasuntojen, hintanousu perustuu muihin ominaisuuksiin. Ne voivat olla esimerkiksi laadultaan, sijainniltaan tai arkkitehtuuriltaan arvostetumpia ja ikämuuttuja toimii näiden tekijöiden proxyna. Korjausrakentamisella voi olla vain ikäkäyrän muutoksia pehmentävä vaikutus.

[14]Ensimmäisellä periodilla lasku 40 ikävuoden jälkeen johtuu yhdestä havainnosta. Uudemman aineiston käyrällä on 17 havaintoa yli 40 vuotiaista rakennuksista.

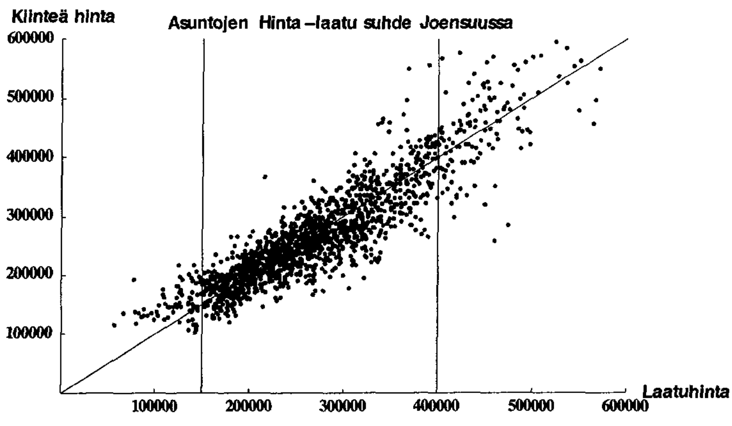


Kuvio 11: Asuntojen laatujakauma Joensuussa. Hedonisten tekijöiden sovitteesta tarkasteluperiodeilta laskettu frekvenssikuvio.

3.3 Asuntojen laatujakauma

Mahdollisten alamarkkinoiden olemassaoloa asuntomarkkinoilla voidaan havainnoida asuntojen laatujakauman kautta. Se saadaan jättämällä pois hedonisen funktion sovitteesta markkinoiden exogeeniset tekijät, tässä tapauksessa ajan vaikutus kauppahintaan ja järjestämällä lasketut arvot suuruusjärjestykseen. Tuloksena on asunnon laadusta ominaisuuksien funktiona kertova "laatuhinta" sovite. Rotherberg t. al. ajatusten mukaan "normaalitasoa" vilkkaampi kaupankäynti jollakin laatutasolla paljastaa itsenäisen alamarkkinan olemassaolon. Kuviossa *11* on esitetty taulukon *3* ja asunnon ikää kuvaavan splini-käyrän mukainen Joensuun asuntojen laatuhinnan jakauma. Siitä ei voi havaita mitään selkeää ympäristöstään erottuvaa kaupankäynnin tihentymää millään laatutasolla. Selvästi suurin osa asunnoista saa laatutasokseen arvon $263,5$ tuhatta ± 50. Pientä keskittymää voisi kuviosta päätellä laatutason $662,5$ ympärille.

Alamarkkinat näkyvät myös markkinahäiriönä. Mikäli asuntojen markkinahinta poikkeaa systemaattisesti laatuhinnasta, ovat näiden asuntojen markkinat eriytyneet omaksi alamarkkinakseen. Nämä tilanteet voidaan havaita helposti asettamalla laatuhinta ja estimoidulla aika-splinillä "deflatoitu" markkinahinta vastakkain samaan kuvioon. Tasapainoisilla markkinoilla tämä suhde muodostaa 45 asteen suoran, mutta systemaattisesti korkeampi tai matalampi hinta tietyllä laatutasolla tulkitaan laatutasoa vastaavaksi alamarkkinaksi. Kuvion *12* mukaan erityisesti asuntojen laatuasteikon alapäässä asunnoista maksetut deflatoidut hinnat pyrkivät ylittämään laatua vastaavan hinnan. Tämä mukaan hyvin vaatimattomat asunnot muodostavat oman alamarkkinansa. Vedämme alamarkkinoiden rajan laatuhinnan tasolle 150 000 mk. Tämän alapuolella asuntojen hinnoittelu pelkästään päämarkkinoiden perusteella avaa mahdollisuuden arbitraasille. Vertailun vuoksi myös hyvin korkealaatuiset asunnot on erotettu kuviossa *12* viivalla. Niille hinta-laatusuhde kyllä vaihtelee voimakkaasti, mutta ei ole vinoutunut ylös- tai alaspäin.



Kuvio 12: Asuntojen laatujakauman indikaattori. Estimoitu laatutaso ja aikatrendistä puhdistetut havainnot.

Kuvion *12* mukainen mallin kyky kuvata markkinoiden toimintaa on toisaalta yllättävän hyvä. Asunnon pinta-alaa, keskustaetäisyyttä ja kerrosta kuvaavat muuttujat ovat vastoin empiiristä kokemusta ja osin teoreettista todistelua estimoitu lineaarisina. Jos näiden muuttujien todellinen vaikutus on aidosti konveksi funktio (vrt Jones [Jones88]) sekä sovitteen ylä- että alapäässä pitäisi ilmetä systemaattista virhettä joka, voitaisiin tulkita virheellisesti alamarkkinaksi. Alamarkkinan paikantuminen pelkästään laatutason alapäähän jättää pienen spesifikaatiovirheen mahdollisuuden. Jätämme tämän mahdollisuuden tarkemmin raportoimatta, sillä mahdollinen spesifikaatiovirhe ei vaikuta löydettyä alamarkkinaa heikentävästi. Liitteessä 2 on kuitenkin kuviota *12* vastaava kuva, joka perustuu asunnon pinta-alan neliöjuureen.

Taulukko 4: Asuntomarkkinoiden jakautuminen alamarkkinoihin

Tuhatta	lukumäärä	%-osuus
Alle 150	74	4.3
150-	1638	95.7
Yhteensä	1712	100

Joensuun asuntomarkkinoiden tehokkuus näkyy siinä, että valtaosalla markkinoista asunnon hinta voidaan johtaa sen ominaisuuksista. Vain hyvin halvoilla asunoilla käydyssä kaupassa hinta on suurempi, kuin mitä laatuhinta antaisi olettaa. Alamarkkinan merkitys näyttäisi taulukon *4* mukaan koko asuntomarkkinoiden kannalta olevan marginaalinen. Ominaisuuksien korvattavuus näyttäisi olevan mahdollista hyvin laajalla laatuhaitarilla ja Joensuulainen asunnonostaja pystyy löytämään markkinoilta hinta-laatu suhteeltaan käyvän asunnon. Kuitenkin taulukko *4* osoittaa vain keskimääräisen tilanteen. Alamarkkinoiden todellinen relevanssi selviää, kun kohdennamme niiden vaikutuksen alueittain.

3.4 Alamarkkinoiden analyysi

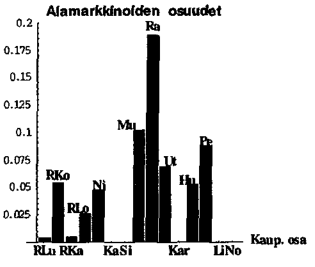
Taulukossa *5* on laskettu keskiarvot ja niiden 95 %:n luottamusvälit kullekin selittävälle muuttujalle. Tarkasteluun on otettu erikseen löydetyn alamarkkinan havainnot

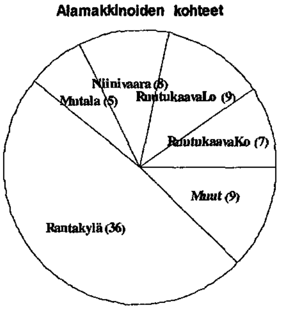
Taulukko 5: Muuttujien keskiarvot ja 95 % luottamusvälit laatuhinnan mukaan ryhmitellyllä aineistolla

Laatuhinta	< 150 000 (alamarkkina)	150 000 - 400 000	> 400 000
PINTA-ALA	$33.11_{[31.27,34.97]}$	$54.58_{[53.67,55.48]}$	$95.12_{[91.04,99.19]}$
KERROS	$2.1_{[1.85,2.35]}$	$2.14_{[2.09,2.18]}$	$2.13_{[1.98,2.28]}$
KUNTO	$1.47_{[1.29,1.67]}$	$1.21_{[1.18,1.24]}$	$1.05_{[1.01,1.1]}$
SAUNA	0	$0.11_{[0.09,0.14]}$	$0.83_{[0.73,0.92]}$
VASTIKE	$8.97_{[8.43,9.5]}$	$8.35_{[8.21,8.47]}$	$6.84_{[6.39,7.28]}$
KESK.ET	$3276_{[2831,3723]}$	$1863_{[1783,1942]}$	$1494_{[1297,1690]}$
RANTA	0	$0.03_{[0.02,0.04]}$	$0.07_{[0.03,0.1]}$
IKÄ	$22.19_{[20.22,24.16]}$	$19.04_{[18.54,19.54]}$	$10.91_{[9.40,12.43]}$
AIKA	$1992.36_{[1991.38,1993.35]}$	$1992.34_{[1992.13,1992.55]}$	$1996.39_{[1991.74,1993.04]}$

sekä silmämääräisesti erotettu markkinoiden laadullinen yläpää (eli laatuhinnaltaan yli 400 000 mk:n asunnot) että "normaalit" asunnot niiden välillä. Lähes kaikki selittävät muuttujat käyttäytyvät laadun mukaisen alamarkkinan suhteen oletetusti. Alamarkkinan asunnot ovat pienempiä, huonokuntoisempia, saunattomia, kalliimpia hoitovastikkeltaan ja kauempana keskustasta ja rannasta ja vanhempia kuin muut myydyt asunnot. Estimoitu malli ei pysty selittämään näistä maksettua korkeaa hintaa. Kaupantekoajankohta (vuositasolla) ei mitenkään vaikuta hinta-laatu suhteeseen. Mielenkiintoista alamarkkinassa on asunnon iän ja kalliin vastikkeen yhdistelmä. Noin 20 vuotiaissa asunnoissa ei yleensä ole rakennusaikaisia yhtiövelkoja maksettavana eikä laajoja peruskorjauksiakaan ole vielä tehty. Tällä ominaisuusyhdistelmällä malli ennustaa laadun heikoksi, mutta todellisuudessa asuntomarkkinat ohittavat tämän tekijän. Korkea vastike voi johtua pienten asuntojen suhteellisesti korkeammasta vastikkeesta.

Asuntojen ikäerot eivät pysty yksin selittämään alamarkkinan syntymistä. Alamarkkinan asunnot ovat vain muutaman vuoden vanhempia kuin laatuhinnalla 150 000-400 000. Kuvion *8* perusteella noin 20 vuoden ikäiselle asunnolle tämä merkitsee vain muutaman tuhannen markan hintaeroa. Asunnon pieni pinta-ala, saunan puutuminen ja huonompi kunto karakterisoivat myös merkitsevästi alamarkkinaa. Näistä saunalla on hyvin suuri merkitys (vrt. taulukko *3*), mutta markkinoiden alapäässä sen todellinen merkitys näyttää vähäisemmältä kuin keskimäärin. Asunnon etäisyys keskustasta karakterisoi myös merkittävästi alamarkkinaa. Tyypillisesti asunto on yli kolmen kilometrin päässä keskustasta—kilometriä kauempana kuin vertailuryhmässä. Tämän kilometrin pitäisi laskea hintaa taulukon *3* mukaan karkeasti 30 000 mk. Joensuussa tälle etäisyydelle asettuvat ainakin Rantakylä, Utra ja Noljakka.





Kuvio 13: Asuinalueiden jakauma alamarkkinalla.

Koska selittävien muuttujien perusteella ei pystytä jäljittämään alamarkkinan syntyä, tarkastellaan seuraavaksi alueen merkitystä alamarkkinan synnyssä.

Tarkastellaan toteutuneita kauppoja eri kaupunginosissa. Kuvion *13* pylväiköissä on Joensuun asemakaavakartan mukaisten kaupunginosien ja alamarkkinoiden jakaumat. Kaupunginosat ovat vasemmalta oikealle: **R**uutukaava-alue (**L**uode, **Ko**illinen, **K**aakko ja **L**ounas), **Ni**inivaara, **Ka**nervala, **Si**ihtala, **Mu**tala, **Ra**ntakylä, **Ut**ra, **Kar**sikko, **Hu**kanhauta, **Pe**nttilä, **Li**nnunlahti ja **No**ljakka.

Suhteellisesti suurin merkitys alamarkkinoilla on Rantakylässä ja Mutalassa, siellä nämä kattavat yli 10 % käydystä kaupasta. Penttilässä kauppojen määrä on hyvin pieni. Alueiden etäisyydellä sinänsä ei näyttäisi olevan ratkaisevaa merkitystä, sillä tarkasteluperiodilla uudisrakennetun Noljakan alueen markkinat näyttäisivät toimineen tehokkaasti. On mahdotonta erottaa tällä aineistolla sitä, onko kyseessä asuntojen kysynnän vai tarjonnan kiristyminen. Kuitenkin toimivilla asuntomarkkinoilla ei ostajan kannalta ole merkitystä sillä, kumman markkinatekijän paineessa hinta muodostuu. Rantakylä ja Mutala tunnetaan Joensuussa halpoina asuntoalueina ja ostaja, jolla ei ole varaa muille markkinoille joutuu alamarkkinalla paradoksaaliseen "köyhällä ei olisi varaa halpaan" tilanteeseen . Tämä kohdentuu erityisesti näiden alueiden pieniin asuntoihin.

Rantakylä suurena asutusalueena kattaa lähes puolet alamarkkinahavainnoista. Kuvion *13* piirakkaa tarkasteltaessa tulee ottaa huomioon aineiston kattavuus, vain noin 40 % kaikesta asuntokaupasta on mukana. Mikäli otoksemme ei ole harhainen, kauppoja on käyty ainakin kaksinkertainen määrä.

Kuvion *13* perusteella markkinahäiriöt saattavat olla alueellisesti merkittäviä. Perinteisesti halpoina pidetyt asuinalueet ovat herkkiä markkinahäiriöille. Tämän tuloksen merkitystä pitäisi arvioida myös asuntopoliittisessa keskustelussa. Jo Joensuun kaltaisessa kaupungissa kysyntäpaine laadultaan vaatimattomia asuntoja kohtaan ylittää tarjonnan. Näiden asuntojen hinnat ovat kohonneet kymmeniä tuhansia yli ominaisuuksien edellyttämän (vrt. kuva *12*) ja maksajina ovat pieniä asuntoja tarvitsevat ensiasunnon ostajat tai sijoittajien kautta asunnon vuokralaiset.

4 Tulosten vertailua perinteisiin menetelmiin

Esittäessämme ranskalaista käyrää ekonomisteille yleensä ensimmäisten kysymysten joukossa on vertailu puhtaasti parametrisiin menetelmiin ja erityisesti dummymuuttujien käyttöön. Mitä ranskalaisella käyrällä saavutetaan suhteessa niihin? Tutkimuksen yksi lähtökohta oli mitata mahdollisia ominaisuuksien korvattavuuden vaihteluja. Tämän suhteen on helppo todeta, että ranskalainen käyrä on joustavampi mittausväline kuin mikään polynominen muoto. Tässä käytetyillä muuttujilla näyttäisi *jälkikäteen* siltä, että molemmat käyrät voitaisiin estimoida polynomisina. Myös dummy-muuttujat ovat tässä suhteessa joustavia, sillä ne eivät sido vaihtelua polynomisesti. Havaintomäärän salliessa kutakin solmua kohden voidaan määritellä dummy, joka kertoo solmuhavainnon kontribuution selitettävään. Jokainen yksittäinen dummy ja koko malli, toisin kuin ranskalainen käyrä, on testattavissa hyvin tunnetussa tilastollisessa kehikossa. Tarkastellaan ensin miten dummy-mallin tulokset eroavat semiparametrisesta.

Yhdistetyssä aineistossa on havaintoja 13 vuodelta ja 49 eri ikäistä asuntoa. Kun jätämme malliin (42) vakiotermin, voimme korvata splini-funktiot 60:llä itsenäisellä dummylla. Jätämme molempien muuttujien ensimmäiset havainnot referenssitasoiksi (nolla-dymmyksi). Yhteenveto estimointituloksesta on taulukossa *6* ja kuviossa *14* on vertailtu dummy-parametreja splineihin.

Taulukko 6: Dummy- ja semiparametrisen mallin vertailutulokset

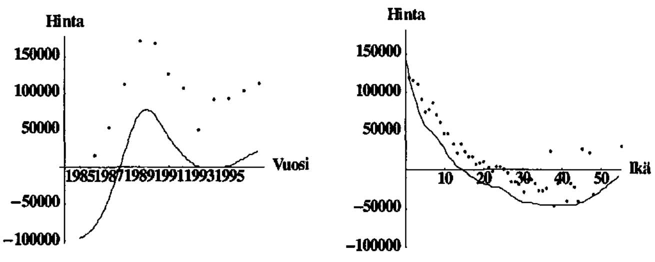
malli	Dymmy	Semiparametrinen
vapausasteet	67	28, 17
RSS	3.29×10^{12}	3.38×10^{12}
F-testisuure	Mallit eivät ole sisäkkäisiä	

Koska mallit eivät ole sisäkkäisiä, ei yksittäisten parametrien tai mallin selityskyvyn muutosta voida testata tavanomaisilla t- ja F- testeillä. Mallien yhteisten lineaaristen muuttujien parametriestimaatit (ei raportoitu) eivät silmämääräisesti arvioituna poikkea toisistaan. Mallien eroa kuvaa hyvin dummy-mallin pienempi residuaalien neliösumma. Dummy-malli estimoi kussakin solmukohdassa keskiarvon muutoin selittämättömästä vaihtelusta. Parametrien arvot määräytyvät simultaanisesti kaikissa solmukohdissa ja havainnossa. Semiparametrisessa mallissa vallitsee sama mekanismi, mutta tämän lisäksi estimoidaan jatkuvat ensimmäiset ja toiset derivaatat solmukohdissa, joten semiparametrinen estimointiongelma on rajoitetumpi kuin dummy-mallilla.

Kuvio *14* kertoo vain estimoiduista ominaisuuksien vaihtelusta, ei vaikutusten tasosta. Olemme lisänneet ikä-dummyyn mallin vakiotermin, ja jättäneet muut käyrät ja estimaatit koskematta. Näin saamme kaikki samaan kuvioon.

Vapaus valita solmukohdan taso ilman kaarevuusrajoitteita näkyy dymmypisteiden voimakkaampana vaihteluna. Erityisesti aikatrendin vuoden 1993 pohjanoteeraus näkyy selvemmin. Jos muutosnopeudet solmukohtien välillä eivät ole kiinnostuksen kohteena, kuten hintaindeksissä usein on kyse, dummy-malli puoltaa paikkaansa paremmin. Toisaalta, jos tutkimuksen kohteena on muutosnopeudet solmu-

Splinifunktiot ja dummy-muuttujat, 1985-1997



Kuvio 14: Splinin ja dymmymuuttujien vertailu. Kutakin dummyä on merkitty pisteellä. Ikä-dummyihin on lisätty mallin vakiotermi siten, että kuviot skaalautuvat samalle tasolle.

kohdissa, kuten voidaan hyvällä syyllä ajatella asunnon iän kohdalla, splini-käyrällä voi estimoida tärkeää informaatiota. Kaikki estimoidut dummy-parametrit olivat tilastollisesti merkitseviä, mutta johtopäätökset asunnon iän vaikutuksesta periodeilla: 5-7, 18-20 tai yli 30 vuotta voisivat olla harhaanjohtavia.

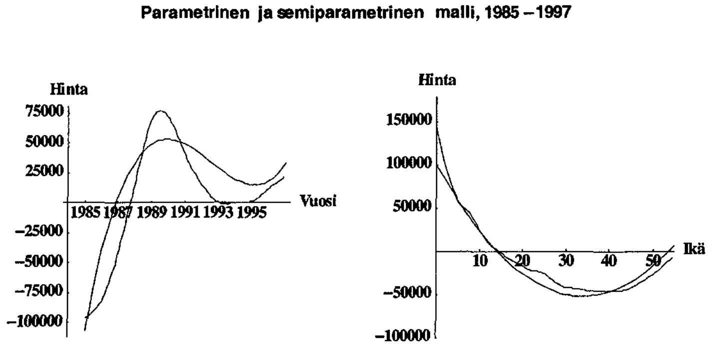
Taulukossa *1* vertasimme semiparametrista mallia lineaariseen. Tämä tarkastelu ei tietenkään tee täyttä oikeutta parametrisille malleille, sillä vakavampi mallitus edellyttää vaihtoehtojen hakemista ja testaamista. Semiparametrinen ja polynominen malli ovat sisäkkäisiä, sillä yhtälössä (2) rajoite $e_i = 0$ tuottaa saman polynomin koko havaintoalueelle. Siten näitä mallivariaatioita on helppo testata tilastollisesti. Taulukossa *7* olemme estimoineet aikatrendille kuutiopolynomin ja asunnon iälle neliön. Kaikki aika- ja ikäpolynomien parametrit ovat tilastollisesti merkitseviä, mutta F-testisuure hylkää parametrisen mallin suhteessa semiparametriseen. Tulos tuntuu yllättävältä aikaisemman tiedon pohjalta, sillä silmämääräisesti polynomien voisi olettaa sopivan hyvin dataan. Kuvio *15* [15] kuitenkin paljastaa aikatrendin huonon sovitteen. Kiinteäparametrisen kuutiopolynomin käyttö tässä sovelluksessa näyttäisi huonolta valinnalta.

Taulukko 7: Polynomi - ja semiparametrisen mallin vertailutulokset.

malli	Polynomi	Semiparametrinen
vapausasteet	12	28, 17
RSS	3.99×10^{12}	3.38×10^{12}
F-testisuure	18.77	

Tämän jälkeen jäisi vielä kokeiltavaksi edellisten hybridi, dummy aikatrendi ja polynominen asunnon ikä vastaan semiparametrinen malli. Tässäkin dummy-termit

[15] Kuviota on jälleen manipuloitu vakiotermin identifikaatio-ongelman vuoksi seuraavasti. Semiparametrisen ja polynomimallin vakiotermien ero on 98185.9. Olemme nostaneet iän polynomiosaa 100000 mk ja laskeneet aikatrendin polynomia 198185.9.



Kuvio 15: Splinin ja polynomin vertailu.

kumoavat mallien sisäkkäisyyden, joten tilastollista päättelyä ei voitaisi suorittaa. Residuaalien neliösumma tulisi kuitenkin tässä mallissa jäämään hyvin lähelle semiparametrista mallia

Vastauksena kysymykseen miksi splinejä pitäisi käyttää, esitämme kolme perusteltua syytä. Ensinäkin ne ovat tehokas ja selkeä tapa visualisoida aineistoa. Toiseksi, mikäli haluamme tietoa hyvin epälineaarisista relaatioista, on splini etukäteistarkastelussa tehokas väline sen havainnointiin. Lopullinen malli voidaan haluttaessa perustaa johonkin yksinkertaistukseen. Kolmanneksi splinillä on mahdollista estimoida muutossuuntia ja -nopeuksia epälineaarisesta tai hyvin poukkoilevasta aineistosta.

5 Johtopäätöksiä

Olemme tutkineet Joensuun asuntomarkkinoiden hintarakenteita ja hintatason muutoksia kauppakohtaisella aineistolla vuosilta 1985 -1997. Tämä ei ole mahdollistanut asuntomarkkinoiden kysyntä- ja tarjontatekijöiden tutkimista, vaan estimoimamme hedoninen hintafunktio kuvaa markkinoilla toteutunutta tilannetta sen jälkeen, kun sekä ostajat että myyjät ovat valintansa tehneet. Kauppahintaa selitetään vain asuntoon tai asuinalueeseen liitettävillä ominaisuuksilla. Eri ominaisuusyhdistelmät kuvautuvat mallissa asunnon laatuhintana, jonka vaihtelusta ja suhteesta markkinahintaan olemme vetäneet johtopäätöksiä asuntomarkkinoiden toiminnasta. Tämän tyyppinen kuvaus hedonisella funktiolla edellyttää hyvin tarkkaa mallin ja muuttujien valintaa. Mallin spesifiointi ja estimointi onkin ollut tutkimuksen toinen keskeinen teema, Joensuun asuntomarkkinoiden hinta-laatu suhteen ohessa.

Asuntojen laadun muutosten mittaaminen edellyttää hyvin tarkkaa asunnon ominaisuuksien ja hinnan sekä ulkopuolisten tekijöiden vaikutuksen erittelyä. Puhtaasti lineaarinen malli tai joku ennalta määrätty polynomi sitoisi vaikutusten suunnan tai niiden muutokset etukäteen katsoen hyvin tiukasti ja tulokset voisivat johtua enemmänkin valitusta funktiomuodosta kuin todellisista markkinoiden ominaisuuksista. Olemme kiertäneet tätä ongelmaa rakentamalla ja estimoimalla semiparametrisen mallin hedoniselle hintafunktiolle. Siinä ne muuttujat, joiden vaikutustapa voidaan olettaa tunnetuksi, otetaan mukaan parametrisina. Ne muuttujat, joiden vaikutustapaa ei etukäteen tunneta tai tiedetään hyvin epälineaariseksi, lisätään malliin ei-parametrisina (tai muuttuvaparametrisina) kuutiosplineinä. Muuttujien ryhmittelyä parametrisiin ja ei-parametrisiin rajoittaa ei-parametrisen osan laskennallinen raskaus. Käytännössä jo kaksi ei-parametrista muuttujaa on haastava suoritus.

Semiparametrisen mallin estimointimenetelmä, roughness penalty, ei ole kovin yleisesti tunnettu ekonomistien keskuudessa, joten olemme uhranneet varsin pitkän osan tekstistä sen seikkaperäiseen kuvaukseen. Estimaattori on lineaarinen, mutta ei-parametrisen osan matemaattinen rakenne ja sitä myötä tarvittava suhteellisen laaja aineisto vaativat paneutumista myös laskentamenetelmiin. Toivomme, että tässä tekemämme työ helpottaa muita soveltajia ja herättää keskustelua ei-parametristen menetelmien soveltuvuudesta erilaisiin tutkimustehtäviin. Estimointimenetelmää pääsee kokeilemaan hakemalla kirjoittamamme Mathematica koodin tämän julkaisun verkkosivulta http://www.vatt.fi. Sen käyttö edellyttää ohjelmapaketin alkeiden hallintaa.

Estimointituloksissa voimme selkeästi hylätä kiinteäparametrisen mallin verrattuna semiparametriseen malliin, jossa asunnon ikä ja aikatrendi ovat mukana luonnollisina kuutiosplineinä. Ero dummy-mallin kanssa ei näytä suurelta, mutta näiden ei-sisäkkäisten mallien testaus ei onnistu perinteisillä tilastollisilla testeillä. Eroja on sen sijaan mallien käyttötarkoituksessa. Dummy-malli voi estimoida vain täysin diskreettejä muutoksia valituissa havaintopisteissä, mutta semiparametrinen malli mahdollistaa muutosnopeuksien estimoinnin koko havaintoalueella. Sen sijaan vertailu (kiinteä)parametrisen mallin kanssa antoi tilastollista tukea semiparametriselle mallille. Tulos on visuaalisesti arvioiden hyvin yllättävä, sillä estimoidut ei-parametriset käyrät rakennuksen iälle ja aika-muuttujalle muistuttavat suuresti paraabelia ja kolmannen asteen polynomia. Kiinteäparametrisen polynomin kuvaaja on kuitenkin aina kulussaan sidottu koko havaintoalueen sovitteeseen, joka jää aina vapaasti polveilevaa ei-parametrista käyrää heikommaksi.

Estimointituloksissa havaitsimme, että asunnon iällä sen myyntihinnan selittäjänä on hyvin vaihteleva merkitys. Rakennuskanta Joensuussa näyttää jakaantuneen kahteen osaan. Uusiin 80- ja 90-luvuilla rakennettuihin asuntoihin iällä on hyvin voimakas merkitys. Ensimmisenä 15 vuotenaan nämä asunnot menettävät hinnastaa karkeasti 100 000 markkaa pelkästään ikätekijän vaikutuksesta. Tähän tulee lisätä vielä yhtiövelkojen maksu sekä muu ikään liittyvä arvon aleneminen, kuten asunnon kunnon heikkeneminen. Toisaalta rakennuskannan vanhalla osalla, ennen 80-lukua rakennetuilla asunnoilla, ikä ei ole enää ratkaiseva tekijä asunnon hinnassa. Asunnot mielletään jo vanhoiksi ja hinta määräytyy paremminkin muiden asunnon ominaisuuksien perusteella. Tälläisellä markkinatilanteella uustuotantoon kohdistuvat hintapaineet jäävät helposti uusasunnon ostajan kannettavaksi, sillä toimivat markkinat painavat asunnon hinnan ajan myötä laatutasoa vastaavaksi. Tällainen ulkopuolinen paine voi syntyä hallituksen talouspoliittisen ministerivaliokunnan toimenpide-ehdotuksesta 31.8.1999, missä kaavoitettujen ja rakentamattomien asuntotonttien verotusta kiristetään. Jos tämä lisää osaltaan painetta rakennuskustannusten ja samalla uusien asuntojen hinnan nousuun, ovat maksajina loppujen lopuksi uusien asuntojen ostajat. He kuluttavat asunnon uutuutta eivätkä pysty siirtämään kohonnutta hintaa asunnon jälleenmyyntiarvoon.

Asuntojen hinta-laatu suhteita tarkasteltiin alamarkkinahypoteesin kannalta. Mikäli markkinoiden hinnoittelu jollakin estimoidulla asuntojen laatutasolla on selvästi yli tai alle laatutason edellyttämän, asuntokauppa tällä laatutasolla muodostaa oman alamarkkinansa. Joensuusta estimoimallamme mallilla havaitsimme, että hyvin vaatimattomista asunnoista maksettiin markkinoilla ylihintaa. Tämän alamarkkinan merkitys on kokonaisuutta katsoen pieni, mutta Rantakylän ja Mutalan alueella 10 - 20 % asuntokaupasta tapahtuu tällä segmentillä. Havainnon tekee poliittisesti ongelmalliseksi se, että markkinoiden kiristyminen kohdistuu halpoja pienasuntoja tarvitseviin joko suoraan asunnon ylihintana ominaisuuksiin nähden tai korkeampana vuokratasona. Ilmeisesti sosiaalinen asuntotuotanto Joensuussa ei pysty näitä markkinoita riittävästi tukemaan.

Joensuu on suhteellisen verkkaisesti kehittyvä alue eikä taloudellinen kasvu viimeisenä 15 vuotena ole kohdistanut mitään merkittävää kysyntäpiikkiä alueen asuntotarjontaan. Silti sosiaalisella asuntotuotannolla voitaisiin tukea markkinoiden toimintaa. Asuntopoliittisen suunnittelun kannalta on tärkeää tarkastella Suomen kasvukeskuksissa sosiaalisen asuntotuotannon mahdollisuuksia tukea markkinoiden toimintaa. Alamarkkinoiden analyysi on tässä yksi varteenotettava mahdollisuus.

Tässä tutkimuksesa on sovellettu uutta estimointimenetelmää, joka sallii tarkastelun ilman rajoittavia oletuksia funktiomuodosta. Sen haittana on kuitenkin jonkin verran perinteistä tilastollista lähetymistapaa vaatimattomammat mahdolliset käyttää tilastollista päättelyä analyysin tukena. Jos näiden tulosten jälkeen on tarvetta tukeutua paremmin tunnettuun parametriseen menetelmään, se voidaan tehdä, jos käytettävä funktiomuoto kykenee approksimoimaan spliniä. Splini-funktion paras käyttöalue onkin aineiston tiivistäminen analyysiä varten. Olemme lisäksi havainnollistaneet, miten sitä voidaan käyttää myös yhtälön estimoinnissa, kun funktion vaihtelua ei haluta etukäteen rajoittaa.

Lähdeluettelo

[BL77] A. Buse and L. Lim. Cubic Splines as a Special Case of Restricted Least Squares. *Journal of the American Statistical Association* **72**:64-68 1977.

[Eub88] Randall L. Eubank. *Spline Smoothing and Nonparametric Regression.* Dekker, New York 1988.

[GS94] P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models.* Chapman & Hall, Bungay 1994. *Errata:* http://www.stats.bris.ac.uk:81/pub/reports/SMOOTH/errata.

[Gre93] William Greene. *Econometric Analysis.* Prentice Hall, New York 1993.

[HT90] T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models.* London 1990.

[Huk90] Sari Hukkanen. *Asuntomarkkinoista ja asuntojen hintojen kehityksestä Joensuussa vuosina 1985-89.* Pro Gradu, Joensuun yliopisto, Kansantaloustiede 1990.

[Joh84] J. Johnston. *Econometric Methods.* McGraw-Hill, 3ed. 1984.

[Jones88] Larry E. Jones. The Characteristics Model, Hedonic Prices, and the Clientele Effect. *Journal of Political Economy* **96**(3):551–567 1988.

[Khu93] André I. Khuri. *Advanced Calculus with Applications in Statistics.* Wiley, New York 1993.

[Kyll98] Lauri Kyllönen. *Asuntomarkkinoiden alamarkkinat - semiparametrinen hedoninen tutkimus Joensuun asuntomarkkinoista.* Pro Gradu, Joensuun yliopisto, taloustieteen laitos 1997.

[Laa97] Seppo Laakso. *Urban Housing Prices and the Demand for Housing Characteristics. A Study on Housing Prices and the Willingness to Pay for Housing Characterstics and Local Public Goods in the Helsinki Metropolitan Area..* **ETLA A**(27), Helsinki 1997.

[MQ96] Carl Mason and John M. Quigley. Nonparametric Hedonic Housing Prices. *Housing Studies* **11**:373–385 1996.

[Mie95] Antti Miettinen. *Hedonisen hintafunktion jäljillä: Ekonometrinen tutkimus vanhojen vapaarahoitteisten kerrostalohuoneistojen hintojen määräytymisestä Joensuussa.* Pro Gradu, Joensuun yliopisto, taloustieteen laitos 1995.

[Pal91] Raymond B. Palmquist. Hedonic Methods. In J.B. Braden and Kolstad C.D.(eds.), *Contributions to Economic Analysis*, pages 77–120, Amsterdam 1991.

[Pea83] Carl E. Pearson. *Handbook of Applied Mathematics : Selected Results Andmethods.* Van Nostrand, New York 1983.

[Phl83] Lois Phlips. *The Economics of Price Discrimination.* Cambridge 1983.

[Poirier73] D. J. Poirier. Piecewise Regression Using Cubic Splines. *Journal of the American Statistical Association* **68**:515-524 1973.

[Ros74] Shrewin Rosen. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* **82**:34–55 1974.

[RGBP91] Jerome Rothenberg, George C. Galster, Richard V. Butler and John R. Pitkin. *The Maze of Urban Housing Markets: Theory , Evidence and Policy.* 1991.

[Sch46a] I. J. Schoenberg. Contributions to the Problem of Approximation of Equidistant Data by Analytic Function. PART A.—On the Problem of Smoothing or Graduation. A First Class of Analytic Approximation Formulae Quarterly of Applied Mathematics **IV**:45-99 1946.

[Sch46b] I. J. Schoenberg. Contributions to the Problem of Approximation of Equidistant Data by Analytic Function. PART B.—On the Problem of Osculatory Interpolation. A Second Class of Analytic Approximation Formulae Quarterly of Applied Mathematics **IV**:112-141 1946.

[WW83] Edward J. Wegman and Wright Ian W. Splines in Statistics. *Journal of the American Statistical Association* **78**(382):351-365 1983.

[Wil81] Oliver E. Williamson. The Economics of Organizations: The Transaction Cost Approach, *American Journal of Sociology*, **87**:548–577 1981

[Yat98] Adonis Yatchew. Nonparametric Regression Techniques in Economics. *Journal of Economic Litterature* **36**(2):669-721 1998.

Tilastolähteet

[Kauhin] Suomen Kiinteistövälittäjien Liitto. *Kauppahintaluettelo 1.1.1985 - 30.4.1993, 1.1.1994 - 28.2. 1997*

[Til] Tilastokeskus. *Asuminen 1990:2, Asuminen 1993:5, Asuminen 1996:5, Asuminen 1997:5* 1990–1997

Liite 1. Parametrirajoitteet PNS estimaattorille

Havainnollistetaan splinin algebrallista rakennetta kolmannen asteen splinillä ja yhdellä solmukohdalla t.

$$P(h) = \begin{cases} a_1 + b_1 h + c_1 h^2 + d_1 h^3, & a \le h \le t \\ a_2 + b_2 h + c_2 h^2 + d_2 h^3, & t \le h \le b \end{cases}$$

Koska funktion arvon sekä ensimmäisen ja toisen derivaatan täytyy olla jatkuvia pisteessä $h = t$

$$\begin{array}{ccc} h \longrightarrow t^- & = & h \longrightarrow t^+ \\ a_1 + b_1 t + c_1 t^2 + d_1 t^3 & = & a_2 + b_2 t + c_2 t^2 + d_2 t^3, \\ b_1 + 2c_1 t + 3d_1 t^2 & = & b_2 + 2c_2 t + 3d_2 t^2, \\ 2c_1 + 6d_1 t & = & 2c_2 + 6d_2 t \end{array}$$

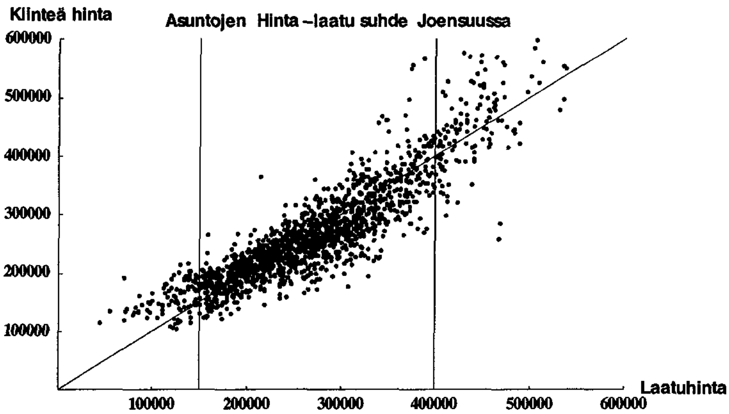
Ratkaisemalla yhtälöryhmää rekursiivisesti voidaan päätellä parametrirajoitteet splinin osien välille:

$$\begin{aligned} a_1 - a_2 &= (d_2 - d_1)\, t^3, \\ b_1 - b_2 &= -3\,(d_2 - d_1)\, t^2, \\ c_1 - c_2 &= 3\,(d_2 - d_1)\, t, \end{aligned} \tag{43}$$

Koska t on valittu solmukohdan arvo, vakio, yhtälöt (43) ovat h:n suhteen lineaarisia rajoitteita. Mikäli käytettävissä on runsaasti havaintoja jokaiselle solmuvälille voidaan havaintomatriisia sopivasti järjestelemällä (katso esimerkiksi Johnston, [Joh84, sivut 392-396]) estimoida splinin polynomit jokaiselle solmuvälille. Toisaalta rajoitettu PNS tuottaisi hyvin epälineaarisen käyrän kuten kappaleesta 2.3 käy ilmi.

Liite 2. Asunnon pinta-alan vaikutus hintaan

Mallin spesifikaatiolla on keskeinen merkitys sovitteelle. Kun kiinteät kustannukset asunnon pinta-alan kasvaessa jyvittyvät isommalle alalle, voidaan hyvin perustein väittää, että asunnon pinta-alalla on laskeva rajavaikutus myyntihintaan. Niinpä neliöjuuri voisi toimia paremmin kuin lineaarinen pinta-ala. Kuviossa *16* on piirretty tätä vaihtoehtoa vastaava asunnon laatuhinnan ja myyntihinnan välinen pisteparvi.



Kuvio 16: Asuntojen laatujakauman indikaattori. Estimoitu laatutaso ja aikatrendistä puhdistetut havainnot.

Samoin kuin kappaleessa 3.3 systemaattiset poikkeamat lävistäjältä kertovat mahdollisista alamarkkinoista. Halpojen ja ylihintaisten markkinoiden osa on edelleen havaittavissa, ja tämän lisäksi kuvion keskivaiheilla hinnat pyrkivät painumaan alle laatutason ja yläpäässä taas ylittämään sen. Nämä lisävaikutukset eivät ole kuitenkaan erityisen selkeitä, joten olemme päätyneet raportoimaan ja tulkitsemaan tuloksia lineaarisen mallin pohjalta.

Mallia voitaisiin vielä varioida jakamalla pinta-alan vaikutus lineaariseen- ja neliöjuurikomponenttiin. Periodilta 1994-1994 estimoituna neliöjuuriosa ei ole tässä vaihtoehdossa tilastollisesti merkitsevä.

Näillä mallivariaatioilla ei ole näkyvää vaikutusta muiden muuttujien kontribuutioon hintamallissamme.

Liite 3. Käytetyt matemaattiset merkinnät

p	Muuttuja p, asunnon velallinen myyntihinta
p_i	i:s havainto muutujasta p
$\mathbf{p}$	Vektoroidut havainnot muuttujasta p
$\mathbf{g}_i$	i:s havaintovektori muuttujatyypistä g
$\mathbf{R}$	Matriisi
t_i	Solmukohta i
T_i	Solmuväli $[t_i, t_{i+1}]$
$P(h), g_i(h)$	Muutujan h splini-funktio, splinin osa solmuvälillä $[t_i, t_{i+1}]$
$g''(h)$	Derivaattamerkintä
$g_i(t_{i+1}), g_i''(t_{i+1})$	$g_i(h)$:n arvo solmussa t_{i+1}, sen toisen derivaatan arvo solmussa t_{i+1}
γ_{i+1}	$g_i''(t_{i+1})$, eli toisen derivaatan arvo solmussa t_{i+1}

VATT-TUTKIMUKSIA -SARJASSA ILMESTYNEITÄ

PUBLISHED VATT-RESEARCH REPORTS

43. Lehtinen Teemu: The Distribution and Redistribution of Income in Finland 1990-1993. Helsinki 1998.

44. Rantala Juha: Työvoimapolitiikan rooli ja työttömien työllistyminen. Helsinki 1998.

45. Laurila Hannu: Suomalaisen kaupunkipolitiikan taloudelliset lähtökohdat. Helsinki 1998.

46. Tuomala Juha: Pitkäaikaistyöttömyys ja työttömien riski syrjäytyä avoimilta työmarkkinoilta. Helsinki 1998.

47. Tossavainen Pekka: Panosverot ja toimialoittainen työllisyys. Helsinki 1998.

48. Holm Pasi – Kiander Jaakko – Tuomala Juha – Valppu Pirkko: Työttömyysvakuutusmaksujen työttömyysriskin mukainen porrastus ja omavastuu. Helsinki 1998.

49. Kari Seppo – Kröger Outi – Rauhanen Timo: Henkilöyhtiöiden verotuksen investointi- ja työllistämiskannustimet. Helsinki 1998.

50. Kajanoja Jouko: Lasten päivähoito investointina. Helsinki 1999.

51. Kari Seppo: Dynamic Behaviour of the Firm Under Dual Income Taxation. Helsinki 1999.

52. Holm Pasi – Sinko Pekka – Tossavainen Pekka: Työpaikkojen syntyminen ja päättyminen ja rakenteellinen työttömyys. Helsinki 1999.

53. Mäkelä Pekka (toim.): EU:n kauppapolitiikkaa itälaajenemisen kynnyksellä. Helsinki 1999.

54. Sinko Pekka: Taxation, Employment and the Environment – General Equilibrium Analysis with Unionised Labour Markets. Helsinki 1999.

55. Rantala Anssi: Finanssikriisit, yritysten nettovarallisuus ja makrotaloudellinen vakaus. Helsinki 1999.

56. Kyyrä Tomi: Post-Unemployment Wages and Economic Incentives to Exit from Unemployment. Helsinki 1999.

57. Korkeamäki Ossi: Yksityisen ja julkisen sektorin palkkoihin vaikuttavat tekijät. Ekonometrinen tutkimus 1987 - 1994. Helsinki 1999.

58. Venetoklis Takis: Process Evaluation of Business Subsidies in Finland. A Quantitative Approach. Helsinki 1999.

59. Kuusi Osmo: Expertise in the Future Use of Generic Technologies – Epistemic and Methodological Considerations Concerning Delphi Studies. Helsinki 1999.

60. Hakola Tuulia: Race for Retirement. Helsinki 1999.

61. Korkeamäki Ossi: Valtion palkat yleisiin työmarkkinoihin verrattuna: vuodet 1989 - 1997. Helsinki 2000.

62. Uusitalo Roope: Paikallinen sopiminen ja yritysten työvoiman kysyntä. Helsinki 2000.

63. Milne David – Niskanen Esko – Verhoef Erik: Operationalisation of Marginal Cost Pricing within Urban Transport. Helsinki 2000.

64. Vaittinen Risto: Eastern Enlargement of the European Union. Transition in applicant countries and evaluation of the economic prospects with a dynamic CGE-model. Helsinki 2000.

65. Häkkinen Iida: Muuttopäätös ja aluevalinta Suomen sisäisessä muuttoliikkeessä. Helsinki 2000.

66. Pyy-Martikainen Marjo: Työhön vai eläkkeelle? Ikääntyvien työttömien valinnat työmarkkinoilla. Helsinki 2000.