

Maanmittaus 81:1-2 (2006)  
Saapunut 10.8.2005 ja tarkistettuna 11.4.2006  
Hyväksytty 30.6.2006

## Maanhintojen vikasietoisesta mallintamisesta

Marko Hannonen

Teknillinen korkeakoulu, Kiinteistöopin laboratorio  
PL 1200, 02015 TKK  
marko.hannonen@tkk.fi

**Tiivistelmä.** Tässä artikkelissa tarkastellaan vikasietoista mallintamismenetelmää, kolmivaiheista MM-estimointia, maanhintojen ekonometrisessä analyysissä. Kyseinen menetelmä on laskentaintensiivinen, jossa 1. vaiheessa lasketaan regressioestimaatti, joka on tarkentuva ja omaa korkean murtumispisteen, mutta joka ei välttämättä ole tehokas. 2. vaiheessa estimoidaan virheiden hajonta, joka perustuu 1. vaiheen jäämöksiin. 3. vaiheessa lasketaan regressioerrointen M-estimaatti soveltaen kaksoispainotettua kohdefunktiota. MM-estimointitekniikka ei ole herkkä poikkeaville tai vaikutusvaltaisille havaintopisteille ja täten kestää hyvin tiettyä määrää epätyypillisiä tai erheellisiä havaintoarvoja. Empiirinen tarkastelu osoitti, että estimoitujen hedonisten maanhintamallien sisäistä tarkkuutta voitiin selvästi parantaa molemmilla analysoiduilla hinta-aineistoilla (Espoo ja Nurmijärvi), kun siirryttiin pienimmän neliösumman keinosta MM-estimointiin. Lisäksi ennustustarkkuutta voitiin kasvattaa pienemmän systemaattisen ennustevirheen muodossa Espoon hinta-aineistolla, kun hintamalli estimoitui MM-tekniikalla.

**Avainsanat:** vikasietoisuus, regressio, maan hinta, poikkeavat havainnot.

### 1 Johdanto

Ekonometriset menetelmät ovat usein välttämättömiä vaativimmissa kiinteistöjen arviointitehtävissä, jotta systemaattinen arviointivirhe voitaisiin objektiivisesti minimoida, ja jotta tarpeelliset hintojen laatukorjaukset voitaisiin määrittää osuvasti ja luotettavasti. Maanhintojen ekonometrisessä mallintamisessa on vähintään viisi merkittävää peruskysymystä. Näitä ovat (1) hintojen ajallinen vaihtelu, (2) hintojen spatiaalinen vaihtelu, (3) mallispesifikaatio-ongelma, (4) poikkeavat ja vaikutusvaltaiset havainnot ja (5) epävarmuuden luonteva kontrolli.

Hintojen aikaulottuvuuden analyysissä on oleellista ymmärtää, että maanhintojen käyttäytyminen on pääsääntöisesti epästationaarista. Tämä on tyypillinen taloudellisiin aikasarjoihin liittyvä ominaispiirre, joka tarkoittaa, että havaintoaineiston tuottamisprosessi (mekanismi, joka tuottaa havaitun hinta-aineiston)

muuttuu ajassa, eikä täten ole vakioprosessi. Ajan aiheuttama hintavaikutus on lisäksi moniulotteinen: Usein voidaan perustellusti erottaa toisistaan pitkän ajanjakson hintakehitys (hintatrendi), hintojen suhdannekierto (hintasykli), kausivaihtelu ja tilapäinen eli satunnainen hintojen vaihtelu. Maanhintojen epästationaarinen luonne ja ajallinen moniulotteisuus tekevät pätevästä hintamallintamisesta erittäin haasteellisen tehtävän. Perinteisesti maanhintojen ajallinen vaihtelu on pyritty palauttamaan elinkustannusindeksin tai asuntojen hintaindeksin arvojen vaihteluun, joita on käytetty hintamallissa aikaa selittävänä tekijänä (ks. esim. Hiltunen 2003). Myös nk. luokkamuuttujatekniikka on ollut suosittu ratkaisuvaihtoehto (usein käytetään vuosiluokkamuuttujia). Näihin kaikkiin liittyy ongelmia ensisijaisesti siksi, että ne eivät välttämättä kykene kuvaamaan todellista ajallista vaihtelua käytännössä riittävällä tarkkuudella. Rakenteelliset aikasarjamallit (ks. Hannonen 2005) ja ns. aalloke-muunnokset eli wavelet-muunnokset (ks. Hannonen 2006a) soveltuvat sen sijaan lähtökohtaisesti huomattavasti paremmin monimutkaiseen epästationaariseen mallintamiseen.

Maanhintojen spatiaalinen vaihtelu voidaan jakaa spatiaaliseen heterogeenisuuteen ja spatiaaliseen riippuvuuteen. Spatiaalinen heterogeenisuus osoittaa, että hintamallin funktionaalinen muoto ja parametrien arvot vaihtelevat sijainnin muuttuessa, kun taas spatiaalinen riippuvuus osoittaa, että hintojen vaihtelu on etäisyyden funktio. Spatiaalinen riippuvuus voidaan usein ratkaista hyväksyttävästi sisällyttämällä sijaintia tai etäisyyttä kuvaavia termejä hintamalliin selittäväksi tekijöiksi. Sen sijaan spatiaalinen heterogeenisuus on usein ongelmallisempi kysymys. Yksi ratkaisuvaihtoehto olisi rajata havaintoaineisto hyvin suppealle osa-alueelle, jolloin tarkoituksena on homogenisoida tutkittavaa ilmiötä riittävässä määrin, mutta tätä toimenpidettä ei ole aina mahdollista luotettavasti toteuttaa havaintoaineiston niukkuuden vuoksi. Joustavat mallintamistekniikat ovatkin usein parempi vaihtoehto, jotka omaavat spatiaalisen adaptaatiokyvyn ja täten eksplisiittisesti huomioivat paikallisen heterogeenisuuden asettamat vaatimukset.

Mallispesifikaatio-ongelmaa voidaan lähestyä kolmella eri tavalla: (1) parametrisesti, (2) semi-parametrisesti ja (3) ei-parametrisesti. Parametrinen mallintaminen on klassisesti sovellettu lähestymistapa kiinteistötalouden ja -arvioinnin ekonometrisissa sovelluksissa (Kantola 1982 ja 1983). Se on luonteeltaan teoria-vetoista ja siinä etukäteen kiinnitetään mallin funktionaalinen rakenne. Ei-parametriset mallintamismenetelmät ovat puolestaan aineistovetoisia, hyvin joustavia tekniikoita, ja semi-parametriset tekniikat yhdistävät ominaispiirteitä sekä parametrisesta että ei-parametrisesta mallintamisesta. Täsmällisestä tutkimusongelmasta riippuu, mitä mallintamisen viitekehikkoa kannattaa soveltaa. Pääsääntöisesti ei-parametriset menetelmät ovat erityisen sopivia työvälineitä silloin, kun muuttujien väliset riippuvuussuhteet ovat monimutkaisia (hyvin epälineaarisia) ja teoreettisesti tuntemattomia. Parametriset mallit taas soveltuvat hyvin vähemmän kompleksisiin mallintamistilanteisiin, joissa on selkeää etukäteistietämystä käytettävästä mallirakenteesta. Lähestymistavasta riippumatta mallispesifikaatio-ongelmassa joudutaan etsimään vastausta hintamallin raken-

neosan funktionaalisen esitysmuodon, mallimuuttujien ja virhejakauman määrittämisestä osalta. Lisäksi aina on huomattava, että tutkimustulokset riippuvat käytetystä – usein implisiittisestä – mittakaavasta.

Poikkeavat ja vaikutusvaltaiset havainnot ovat hyvin tyypillisiä maanhinta-aineistoissa, jotka saattavat olla aitoja, virheettömiä havaintopisteitä, mutta joiden olosuhteisiin liittyy jotain epätyypillisiä tekijöitä, jotka heikentävät havaintoaineiston edustavuutta, tai ne voivat sisältää erilaisia virheitä (kuten tallennus- ja mittausvirheitä, väärä osajoukko, jne.). Perinteiset mallintamistekniikat, erityisesti tavallisesti sovellettu klassinen pienimmän neliösumman keino, ovat hyvin herkkiä poikkeaville havaintopisteille; jopa yksittäinen poikkeava havaintopiste saattaa oleellisesti muuttaa tuloksia ja vääristää tulkintaa. Itse asiassa yksittäinen riittävän poikkeava havaintopiste riittää, että pienimmän neliösumman estimaattori murtuu täydellisesti tuottaen tulkinnallisesti epäluotettavia ja epäinformatiivisia tuloksia. Vikasietoiset (robustit) työvälineet eivät sen sijaan ole herkkiä poikkeaville tai vaikutusvaltaisille havaintopisteillä ja täten kestävät hyvin tiettyä määrää epätyypillisiä tai erheellisiä havaintoarvoja ilman, että estimaattori murtuu ja tuottaa virheellisiä tuloksia.

Epävarmuuden laskenta ja tämän upottaminen luonnollisella tavalla osaksi analyysia on tärkeä tilastollinen aspekti, sillä talousagenttien toiminta maamarkkinoilla voidaan yleisesti ymmärtää päätöksenteoksi epävarmuuden vallitessa. Epävarmuutta (perinteisen satunnaisuuden sijasta) voidaan kontrolloida nk. bayesiläisten menetelmien avulla. Teoriassa harhaton ja tehokas maanhintamalli pystyy huomioimaan legitimoitusti tai ainakin likimääräisesti kaikki edellä mainitut viisi osaongelmaa. Käytännön mallintamistilanteissa on kuitenkin usein vaikeata samanaikaisesti huomioida kaikkia viittä ongelma-aluetta ja joitakin analyysia yksinkertaistavia lähtöolettamuksia joudutaan pakostakin kiinnittämään. Vikasietoiset työvälineet ovat tällaisissa tilanteissa, joissa mallintamisen lähtöolettamukset eivät täysin toteudu vaan ovat ainoastaan likimääräisesti voimassa, erittäin varteenotettava viitekehikko maanhintoja tutkittaessa hedonisen hintateorian puitteissa ekonometrisesti. Perussääntönä on, että hedonisen hintateorian mukaiset lähtöolettamukset ovat aina tietyssä määrin idealisoituja eivätkä sellaisenaan toteudu käytännön ekonometrisessä maanhintatutkimuksessa.

### ***1.1 Tutkimusongelma ja metodiikka***

Tässä artikkelissa keskitytään maanhintojen vikasietoiseen (eli robustiin) parametrisen ekonometrisen mallintamiseen; tutkimusongelma on spesifinen tarkasteltavien osamarkkinoiden ja maankäyttölajin osalta, mutta käytetty analyttinen metodiikka on yleinen ja sovellettavissa laajemminkin kiinteistömarkkinoilla. Hieman toisin ilmaistuna tässä artikkelissa yhtäältä keskitytään poikkeavien ja vaikutusvaltaisten havaintopisteiden problematiikkaan hintamalleja parametrisesti estimoitaessa ja sovellettaessa maamarkkinoilla sekä toisaalta paneudutaan tulkinta- ja päätöksentekotilanteisiin maamarkkinoilla, kun hedonisen hintateorian lähtöolettamukset ovat ainoastaan likimääräisesti paikkansa pitäviä. Artikkelissa tarkastellaan parametrissa erittäin vikasietoista maanhinto-

jen mallintamista soveltamalla kolmivaiheista MM-estimointitekniikkaa (ks. Yohai 1987; Marazzi 1993, s. 201). Sovellettavan menetelmän nk. murtumispiste (piste, jossa estimaattori murtuu tuottaen hyödyttömiä tuloksia) on korkein mahdollinen eli 50 %, ts. puolet havaintoaineistosta voi olla ”pilaantunutta” ennen kuin estimaattori tuottaa täysin hyödyttömiä tuloksia. Hieman täsmällisemmin ilmaistuna estimaattorin murtumispisteellä tarkoitetaan sitä havaintoaineiston määrää, joka voi saada äärettömän suuria arvoja ja kuitenkin samanaikaisesti estimaattori säilyy arvoiltaan rajoitettuna. Näin ollen estimaattorin murtumispiste on luonnollinen vikasietoisuuden mittari. Kyseinen menetelmä on hyvin laskentaintensiivinen, jossa 1. vaiheessa lasketaan regressioestimaatti, joka on tarkentuva ja omaa korkean murtumispisteen, mutta joka ei välttämättä ole tehokas. 2. vaiheessa estimoidaan virheiden hajonta, joka perustuu 1. vaiheen residuaaleihin (jännöksiin). 3. vaiheessa lasketaan regressiokerrointen M-estimaatti soveltaen kaksoispainotettua kohdefunktiota eli lasketaan uusi regressiokertoimien vektori painotetuin havainnoin.

Maanhintamallintamisen lähestymistapa on tässä artikkelissa parametrinen (ja ei-spatiaalinen): Maanhintamallin rakenneosan funktionaaliseksi esitysmuodoksi on valittu tavallinen tulomuotoinen joustomalli, jonka oletetaan huomioivan muuttujien välisen epälineaarisen käyttäytymisen riittävällä tarkkuudella. Toisin sanoen maanhintojen mallirakenteeksi on valittu

$$y_i = e^{\beta_0} \prod_{j=1}^k x_{ij}^{\beta_j} \prod_{j=1}^s e^{\gamma_j d_{ij}} e^{\varepsilon_i} = e^{\beta_0} x_{i1}^{\beta_1} x_{i2}^{\beta_2} \dots x_{ik}^{\beta_k} e^{\sum_{j=1}^s \gamma_j d_{ij}} e^{\varepsilon_i} \quad \forall i \in n. \quad (1)$$

missä on  $k$  kappaletta jatkuvia muuttujia  $x_{ij}$  ja  $s$  kappaletta luokkamuuttujia  $d_{ij}$ .  $\beta_j$  ja  $\gamma_j$  ovat tuntemattomia hedonisia hintoja eli tuntemattomia estimoitavia regressiokertoimia. Empiirinen Box-Cox -muunnos tukee myös tätä mallirakennetta. Pelkistävää normaalijakaumaoletusta ei ole tarpeen vikasietoisessa mallintamisessa tehdä virheiden osalta; virhejakauma voikin olla yhtäältä huipukas ja toisaalta vino. Hieman täsmällisemmin ilmaistuna virhetermit  $\varepsilon_i$  oletetaan nyt riippumattomasti ja identtisesti jakautuneiksi satunnaismuuttujiksi, joiden odotusarvo on nolla eli  $E(\varepsilon_i) = 0$ . Tavallista tulomuotoista hintamallirakennetta on sovellettu yleisesti maanhintatutkimuksissa (ks. Hiltunen 2003), koska tällä on lukuisia etuja. Se on ensinnäkin yksinkertainen (estimoinnin kannalta lineaarinen malli), vähäparametrinen (toteuttaa Occamin partaveitsi -periaatteen) ja helposti tulkittava mallirakenne (estimoidut hedoniset hinnat ovat vakioisia hintajoustoja, joilla on selkeä taloudellinen tulkinta). Toiseksi se huomioi rakenteessaan mallimuuttujien moniulotteisen, epälineaarisen yhdysvaikutuksen. Kolmanneksi estimointia varten siinä skaalataan muuttujat siten, että estimoidut hedoniset hinnat ovat riippumattomia käytetystä mittayksiköstä. Neljänneksi se sallii perusmuodossaan heteroskedastisen virhevarianssin tapauksen. Lisäksi mallin riippuvuus-suhteet ovat tyypillisesti helposti linearisoitavissa kaksoislogaritimuunnoksella, jossa luonnollinen logaritmi otetaan sekä tutkittavan muuttujan että mitta-asteikollisten selittäjien havaintoarvoista, ja jonka jälkeen lineaaristen mallien estimointiteoriaa voidaan tehokkaasti hyödyntää.

## 1.2 Aikaisempi tutkimus

Vikasietoisia regressiotekniikoita on sovellettu yllättävän vähän kiinteistömarkkinoilla ja laajemminkin taloustieteissä. Syinä tähän ovat olleet (Zaman ym. 2001):

- virheellinen uskomus, että suuri otoskoko tekee vikasietoisista työvälaineistä tarpeettomia,
- uskomus, että poikkeavat havainnot voidaan karsia pois aineistosta visuaalisesti tai käyttämällä pienimmän neliösumman residuaaleja tähän tarkoitukseen,
- useiden erityyppisten vikasietoisten tekniikoiden olemassaolo ilman selkeää ohjeistusta, mikä näistä menetelmistä on tarkoituksenmukaisin,
- tottumattomuus vikasietoisen analyysin tulosten tulkinnassa,
- tietämättömyys vikasietoisten työvälaineiden hyödyistä todellisilla aineistoilla.

Kiinteistömarkkinoiden alan julkaisuissa löytyi ainoastaan kaksi selkeää tutkimusta, joissa keskityttiin vikasietoisten menetelmien tutkimiseen kiinteistömarkkinoiden ilmiöiden mallinnuksessa.

Kiinteistömarkkinoilla Janssen ym. (2001) sovelsivat vikasietoista pienimmän neliömediaanin menetelmää, jossa minimoidaan jäännösten neliöiden mediaani perinteisen jäännöseliösumman sijasta, Tukholman kantakaupungissa vuosina 1992–1994 myytyjen asuinkiinteistöjen hinta- ja tuottosuhteiden määrittämiseen. Nimenomaisesti bruttopääomitusaste identifioitiin käyttäen tavallista tulomuotoista joustomallia ja sovelletun pienimmän neliömediaanin tehokkuutta pyrittiin parantamaan yksivaiheisella uudelleen painotetun pienimmän neliösumman keinolla. Johtopäätöksenä he totesivat, että heidän käyttämänsä regressiot ovat luotettavia sekä tarkempia kuin perinteisen pienimmän neliösumman keinon estimaatit.

Thorson (1994) tutki pienimmän neliömediaanin keinoa poikkeavien havaintojen identifiointiin Illinoisissa vuosina 1979–1987 tehtyjen maa-aluekauppojen hinta-aineistossa. Johtopäätöksenä hän totesi, että muutamalla pienimmän neliömediaanin keinon mielessä poikkeavalla havaintopisteellä (standardoitu jäännös on suurempi kuin 2,5) on merkittävä vaikutus estimoinnin tulokseen: Estimoitujen regressiokerrointen suuruus ja myös suunta muuttuivat useassa tapauksessa, kun poikkeavien havaintojen vaikutusta arvioitiin. Käytetty mallirakenne oli puolilogaritminen hintamalli.

Klassisen pienemmän neliösumman mukaisia maanhintatutkimuksia on Suomen markkinoilla julkaistu useita, jotka ovat myös relevantteja lähdeviitteitä vikasietoisessa mallintamisessa. Näitä tutkimuksia ovat mm. Leväisen (1991), Heinosen (1993) ja Hiltusen (2003) maanhintatutkimukset, joiden perusteella voidaan todeta, että tonttien tärkeimpiä hintatekijöitä ovat keskustaetäisyys, rakennusoikeus, tonttitehokkuus, pinta-ala, alueen asukasluku, alueen tulotaso, sijainti suhteessa vesistöön, tonttilaji ja kaupanteon ajankohta. Painotetun pienimmän neliösumman keinoa maanhinnan mallintamisessa Suomen markkinoilla on tutkittu mm. Hannosen (2006b) artikkelissa, jossa hintamalleja estimointiin

ko. tekniikalla joustavasti eli ei-parametrisesti, ja jossa tuloksien vikasietoisuutta parannettiin toisessa vaiheessa hyödyntämällä M-estimoinnin muunnelmaa. Hannosen (2006b) artikkelissa käytetty tutkimusmenetelmä on siis ei-parametrinen vikasietoinen estimointimenetelmä. Yhteenvedona tästä tutkimuksesta voidaan todeta, että joustava ja vikasietoinen mallintaminen johti selvästi elinvoimaisempiin, sekä selitysvoimakkuudeltaan että ennustuskyvyltään parempiin malleihin toisella käytetyistä hinta-aineistoista. Painotetun pienimmän neliösumman menetelmää on tutkittu myös mm. McMillenin (1996) ja Wallacen (1996) artikkeleissa.

### 1.3 Vikasietoisesta mallintamisesta

Vikasietoisen tilastotieteen päämääränä on tutkia (parametristen) estimaattorien käyttäytymistä, kun mallintamisen lähtöolettamukset (lineaarisuus, riippumattomuus, normalisuus, jne.) eivät pidä täsmällisesti paikkaansa vaan ovat voimassa ainoastaan likimääräisesti. Hieman tarkemmin ilmaistuna, vikasietoisen mallintamisen perustavoitteita ovat (Hampel ym. 1986, s. 11):

- määrittää mallirakenne, joka sovituu optimaalisesti pääosaan aineistosta,
- identifioida poikkeavat havainnot (outlierit) ja poikkeavat havaintoaineiston osarakenteet tarkempaa analyysia varten,
- identifioida ja varoittaa mahdollisista vaikutusvaltaisista havaintopisteistä (nk. vipupisteistä),
- ottaa huomioon mahdolliset poikkeamat oletetuista virheiden korrelaatiorakenteista.

Käytännössä tilastollisten mallien likimääräinen luonne on pitkälti seurausta karkeiden virheiden olemassaolosta, mallien empiirisestä luonteesta eli aineistovetoisuudesta sekä teoreettisten oletusten vain osittaisesta paikkansapitävyydestä. Erityisesti kiinteistömarkkinoilla sijainti on tekijä, jossa ei voida tyhjentävästi määritellä kaikkia kysyntä- ja tarjontapuolen vaikutuksia erikseen.

Vikasietoinen tilastotiede voidaan määritellä ja ymmärtää ainakin kahdella eri tavalla. Yhtäältä se tutkii sellaisia tilastollisia teorioita, jotka liittyvät poikkeamiin idealisoiduista mallintamisen lähtöolettamuksista (Huber 2004, s. 1; Hampel ym. 1986, s. 6–7) ja toisaalta hieman rajoitetummin vikasietoisuus voidaan käsittää likimääräisten parametristen mallien tilastotieteeksi. Oleellista molemmista lähestymistavoissa on se, että vikasietoiset työvälineet korvaavat oletetun (parametrisen) mallin tämän täydellisellä ympäristöllä.

Tilastollinen estimointimenetelmä on vikasietoinen, jos (ks. Huber 2004, kappale 1):

- se on kohtuullisen harhaton ja tehokas,
- pienet poikkeamat mallintamisen lähtöoletuksista eivät heikennä tilastollisen mallin suorituskykyä merkittävästi,
- suuremmat poikkeamat eivät kokonaan tuhoa mallin elinvoimaisuutta.

## 2 MM-estimointi

### 2.1 Lineaarinen regressiomalli

Tässä osiossa käydään läpi artikkelissa sovellettavan MM-estimoinnin kannalta keskeiset asiat. Tarkastellaan tavallista parametrista lineaarista regressiomallia:

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \quad 1 \leq i \leq n, \quad (2)$$

missä havaitut hinnat  $y_i \in \mathfrak{R}$ , ominaisuusvektori  $\mathbf{x}_i \in \mathfrak{R}^p$  ja  $\boldsymbol{\beta} \in \mathfrak{R}^p$  on tuntemattomien regressiokertoimien vektori (nk. hedoninen hintavektori).  $\varepsilon_i$  on riippumattomasti ja identtisesti jakautunut stokastinen virhetermi. Pienimmän neliösumman estimaatti määritellään parametrin  $\hat{\boldsymbol{\beta}}$  arvoksi, joka minimoi jäännösumman:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n r_i^2(\boldsymbol{\beta}), \quad (3)$$

missä havaitut virheet  $r_i(\boldsymbol{\beta})$  määritellään:

$$r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{\beta}'\mathbf{x}_i. \quad (4)$$

Kun virheet ovat normaalijakautuneita, niin  $\hat{\boldsymbol{\beta}}$  on suurimman uskottavuuden estimaatti. Tällöin  $\hat{\boldsymbol{\beta}}$  on tehokas.

### 2.2 MM-estimaatit

Hajonnan (mittakaavan) M-estimaatit voidaan määrittellä (Huber 2004; Yohai 1987) seuraavasti. Olkoon  $\rho$  reaaliarvoinen funktio, joka toteuttaa seuraavat lähtöolettamukset:

- (i)  $\rho(0) = 0$  eli funktio saa arvon 0 origossa.
- (ii)  $\rho(-\varepsilon) = \rho(\varepsilon)$  eli funktio on parillinen.
- (iii)  $0 \leq \varepsilon \leq v \Rightarrow \rho(\varepsilon) \leq \rho(v)$  eli funktio on kasvava.
- (iv)  $\rho$  on jatkuva.
- (v) Olkoon  $a = \sup \rho(\varepsilon)$ , jolloin  $0 < a < \infty$ .
- (vi) Jos  $\rho(\varepsilon) < a$  ja  $0 \leq \varepsilon < v$ , niin  $\rho(\varepsilon) < \rho(v)$ .

Hajonnan estimaatti  $s(\boldsymbol{\varepsilon})$  määritellään  $s$ :n arvoksi, joka on ratkaisu yhtälöön:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{\varepsilon_i}{s}\right) = b, \quad (5)$$

missä  $b$  voidaan määrittellä  $E_\phi(\rho(\varepsilon)) = b$  ja missä  $\phi$  kuvaa standardinormaalijakaumaa. Nyt MM-estimaatti voidaan määrittellä kolmivaiheisesti seuraavasti:

1. Laske tarkentuva ja korkean murtumispisteen (mahdollisesti 0,5) omaava estimaatti  $\hat{\boldsymbol{\beta}}_0$ . Estimaatin ei tarvitse olla tehokas.

2. Laske virheet  $r_i(\hat{\boldsymbol{\beta}}_0) = y_i - \hat{\boldsymbol{\beta}}_0' \mathbf{x}_i$ ,  $1 \leq i \leq n$ , ja laske kaavan (5) määritteleminen M-hajonta  $s(\mathbf{r}(\hat{\boldsymbol{\beta}}_0))$  käyttäen vakiota  $b$  siten, että  $b/a = 0,5$ , missä  $a = \max \rho_0(\varepsilon)$  (ja missä hajonnan estimoinnissa käytetty funktio on  $\rho_0$ ).

3. Olkoon  $\rho_1$  toinen funktio, joka toteuttaa lähtöoletukset (i)–(vi) ja että  $\rho_1(\varepsilon) \leq \rho_0(\varepsilon)$  sekä  $\sup \rho_1(\varepsilon) = \sup \rho_0(\varepsilon) = a$ . Olkoon  $\psi_1 = \rho_1'$ . Tällöin MM-estimaatti  $\hat{\boldsymbol{\beta}}_1$  on mikä tahansa ratkaisu yhtälöön:

$$\sum_{i=1}^n \psi_1 \left( r_i(\boldsymbol{\beta}) / s \right) \mathbf{x}_i = 0, \quad (6)$$

joka toteuttaa:

$$S(\hat{\boldsymbol{\beta}}_1) \leq S(\hat{\boldsymbol{\beta}}_0), \quad (7)$$

missä:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_1 \left( r_i(\boldsymbol{\beta}) / s \right) \quad (8)$$

ja missä  $\rho_1(0) = 0$ .

Tässä tutkimuksessa  $\rho_1$ :ksi on valittu kaksoispainotettu kohdefunktio (ks. esim. Fox 1997, s. 408):

$$\rho_1 = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[ 1 - \left( \frac{r}{k} \right)^2 \right]^3 \right\} & , \quad \text{kun } |r| \leq k \\ \frac{k^2}{6} & , \quad \text{kun } |r| > k \end{cases}, \quad (9)$$

missä  $k$  on nk. hienosäätövakio. Mitä pienempi  $k$  on, sitä vikasietoisempia estimoidut hedoniset hinnat ovat. MM-estimaatti siis saadaan, kun 1. vaiheessa lasketaan regressioestimaatti, joka on tarkentuva ja omaa korkean murtumispisteen, mutta joka ei välttämättä ole tehokas. 2. vaiheessa estimoidaan virheiden hajonta, joka perustuu 1. vaiheen jäännöksiin. 3. vaiheessa lasketaan regressioerrointen M-estimaatti soveltaen kaksoispainotettua kohdefunktiota eli lasketaan uusi parametrien kerroinvektori painotetuin havainnoin.

### 2.3 Laskenta-algoritmi

Tässä tutkimuksessa käytettävä MM-estimaatin laskenta-algoritmi on muunnos M-estimaatin laskentaan sovellettavasta iteroituvasta uudelleenpainotetusta pienimmän neliösumman keinosta. Iteroituvaa ratkaisuvaihtoehtoa tarvitaan, koska painot riippuvat residuaaleista, residuaalit riippuvat estimoiduista hedonista hin-

noista ja estimoidut hedoniset hinnat riippuvat painoista. Oletetaan, että meillä on käytettävissä alkuestimaatti  $\hat{\boldsymbol{\beta}}_0$  ja tähän liittyvän hajonnan estimaatti  $s$ . (Vaiheet 1 ja 2 on ratkaistu soveltaen nk. S-estimointia (ks. Rousseeuw ja Yohai 1984). 1., 2. ja 3. vaiheen estimointi on suoritettu käyttäen tilastollista ilmaisohjelmaa R.) Määritellään kaikille  $\boldsymbol{\beta} \in \mathfrak{R}^p$  painot:

$$w_i(\boldsymbol{\beta}) = \frac{\psi_1\left(\frac{r_i(\boldsymbol{\beta})}{s}\right)}{\frac{r_i(\boldsymbol{\beta})}{s}}. \quad (10)$$

Määritellään seuraavaksi:

$$\mathbf{g}(\boldsymbol{\beta}) = \frac{1}{s^2} \sum_{i=1}^n w_i(\boldsymbol{\beta}) r_i(\boldsymbol{\beta}) \mathbf{x}_i = \frac{1}{s} \sum_{i=1}^n \psi_1\left(\frac{r_i(\boldsymbol{\beta})}{s}\right) \mathbf{x}_i \quad (11)$$

ja:

$$\mathbf{M}(\boldsymbol{\beta}) = \frac{1}{s^2} \sum_{i=1}^n w_i(\boldsymbol{\beta}) \mathbf{x}_i \mathbf{x}_i'. \quad (12)$$

-  $\mathbf{g}(\boldsymbol{\beta})$  on  $S(\boldsymbol{\beta})$ :n gradientti. Tarpeelliset iteraatiokaavat voidaan nyt kirjoittaa muodossa:

$$\boldsymbol{\beta}_{j+1} = \boldsymbol{\beta}_j + \frac{1}{2^k} \Delta(\boldsymbol{\beta}_j), \quad (13)$$

missä

$$\Delta(\boldsymbol{\beta}) = \mathbf{M}^{-1}(\boldsymbol{\beta}) \mathbf{g}(\boldsymbol{\beta}) \quad (14)$$

ja missä kokonaisluku  $k$  valitaan siten, että vasen puoli epäyhtälöstä:

$$S\left(\boldsymbol{\beta}_j + \frac{1}{2^k} \Delta(\boldsymbol{\beta}_j)\right) \leq S(\boldsymbol{\beta}_j) - \delta \left(\frac{1}{2^k} \Delta(\boldsymbol{\beta}_j)\right)' \mathbf{g}(\boldsymbol{\beta}_j) \quad (15)$$

on minimissään ja  $0 < \delta < 1$ .

### 3 Hintamallien empiirinen estimointi

#### 3.1 Havaintoaineisto

Empiiristen maanhintatutkimuksien tulokset ja johtopäätökset ovat tunnetusti herkkiä havaintoaineiston muutoksille. Valtaosa kyseisestä herkkyydestä selittyy hintojen ajallis-paikallisilla muutoksilla: mallirakenne ja parametrit muuttuvat sijainnin funktiona ja eivät täten ole homogeenisiä läpi koko havaintoaineiston, kun taas ajallisesti muuttuvat markkinatilanteet aiheuttavat, että havaintoaineiston tuottamisprosessi on epästationaarinen. Lisäksi tulosten herkkyyttä lisäävät tyypillisesti poikkeavat ja vaikutusvaltaiset havainnot, joita on tunnetusti vaikea tyydyttävästi käsitellä perinteisin menetelmin. Otosriippuvuuden vähentämiseksi tässä tutkimuksessa tarkastellaan kahta erillistä havaintoaineistoa, jotka sijaitse-

vat eri osamarkkinoilla, ja jotka liittyvät osittain toisistaan poikkeaviin ajanjaksoihin. Maankäyttötyyppi on sen sijaan kiinnitetty, jotta tarpeetonta maanhinta-aineiston heterogeenisuutta voitaisiin vähentää. Maankäyttötyyppi edustaa detaljikaava-alueen ulkopuolisia, rakentamattomia rakennuspaikkoja, jotka eivät ole saavuttaneet korkeinta ja parasta käyttötarkoitusta, ja jotka on osoitettu yleiskaavassa asuinrakennustarkoituksiin.

Ensimmäinen maanhinta-aineisto liittyy Espoon osamarkkinoihin, jossa hinta- ja ominaisuustiedot ovat kerätty vuosien 1990–2001 väliseltä ajanjaksolta kokonaistutkimuksena (eli mitään varsinaista otantaa ei ole suoritettu). Havaintoaineiston koko on 400 havaintoja, joista viimeisen vuoden 2001 havainnot (yhteensä 39 kappaletta) on jätetty syrjään estimoinnista ennustustutkimusta varten. Laadullisesti kyseinen hinta-aineisto on korkeatasoinen, sillä se on tarkastettu etukäteen virheitä varten ja analysoitu aikaisemmin hintamallitutkimuksessa. Taulukossa 1 on esitetty tyypillisiä otostunnuslukuja Espoon havaintoaineiston muuttujille.

Toinen maanhinta-aineisto liittyy Nurmijärven kunnan osamarkkinoihin, missä kauppahinta-aineiston hinta- ja ominaisuustiedot ovat kerätty lähes 20 vuodelta alkaen 1. tammikuuta 1985 ja päättyen maaliskuuhun 2004. Havaintoaineiston koko on yhteensä 793 havaintoa. Nurmijärven kunnan hinta-aineistossa on kolme selvää osakeskusta (kirkonkylä, Klaukkala ja Rajamäki), joihin kuhunkin on mitattu suoraviivaiset etäisyydet Helsinki-etäisyyden ohella. Viimeisen (noin) vuoden havainnot, joita on yhteensä 50 kappaletta, on jätetty syrjään estimoinnista ennustustarkasteluja varten. Laadullisesti tämä hinta-aineisto ei ole yhtä korkeatasoinen kuin Espoon hinta-aineisto, sillä sitä ei ole erityisesti tarkastettu ekonometrista maanhinta-analyysia varten etukäteen. Taulukossa 2 on esitetty tavallisia otostunnuslukuja Nurmijärven havaintoaineiston muuttujille.

**Taulukko 1.** Tavallisia otostunnuslukuja Espoon hinta-aineistossa.

<b>Muuttuja (yksikkö)</b>	<b>Aritmeettinen keskiarvo</b>	<b>Minimiarvo</b>	<b>Maksimiarvo</b>	<b>Otoshajonta</b>
Kokonaishinta (€)	59 126,40	3 027,00	756 846,00	61 976,88
Neliöhinta (€/m <sup>2</sup> )	22,99	0,24	127,55	19,09
Pinta-ala (m <sup>2</sup> )	4 207,49	1 000,00	28 400,00	4 613,75
Etäisyys Helsingin ydinkeskustaan (km)	17,22	7,61	27,29	4,34
Omakotikiinteistöjen neljännesvuosittainen hintaindeksi (pääkaupunkiseutu)	154,06	116,80	187,30	22,35
Koko (kiinteistö = 0 / määräala = 1) luokkamuuttuja;	-	0	1	-
luokkamuuttujan %-osuus		58	42	

**Taulukko 2.** Tavallisia otostunnuslukuja Nurmijärven hinta-aineistossa.

<b>Muuttuja (yksikkö)</b>	<b>Aritmeettinen keskiarvo</b>	<b>Minimiarvo</b>	<b>Maksimiarvo</b>	<b>Otoshajonta</b>
Kokonaishinta (€)	22 019,03	673,00	479 336,00	21 262,25
Neliöhinta (€/m <sup>2</sup> )	3,80	0,34	22,83	3,27
Pinta-ala (m <sup>2</sup> )	7 387,36	1 000,00	30 000,00	4 651,66
Etäisyys Helsingin ydinkeskustaan (km)	33,11	22,28	44,91	5,22
Etäisyys Nurmijärven kirkonkylään (km)	8,55	0,32	16,06	3,37
Etäisyys Klaukkalaan (km)	9,33	0,50	21,44	5,00
Etäisyys Rajamäkeen (km)	11,86	0,43	20,92	4,58
Omakotikiinteistöjen neljännesvuosittainen hintaindeksi (pääkaupunkiseutu)	154,30	100,00	226,00	32,00
Koko (kiinteistö = 0 / määräala = 1) luokkamuuttuja;	-	0	1	-
luokkamuuttujan %-osuus		36	64	

**Taulukko 3.** Pienimmän neliösumman regression tulokset (Espoo; poikkeavat havainnot mukana).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Pinta-ala	-0,79	0,05	-15,90	0,00
Etäisyys Helsingin ydinkeskustaan	-1,25	0,17	-7,51	0,00
Asuntojen hinta-indeksi	2,52	0,10	25,70	0,00
Koko kiinteistö / määräala dummy	-0,37	0,08	-4,71	0,00
Selitysaste (keskistämätön malli)			0,58	
Regression keskivirhe			0,71	

**Taulukko 4.** Pienimmän neliösumman regression tulokset (Espoo; poikkeavat havainnot poistettu).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Pinta-ala	-0,75	0,04	-18,90	0,00
Etäisyys Helsingin ydinkeskustaan	-1,38	0,13	-10,40	0,00
Asuntojen hinta-indeksi	2,54	0,08	32,70	0,00
Koko kiinteistö / määräala dummy	-0,30	0,06	-4,83	0,00
Selitysaste (keskistämätön malli)			0,68	
Regression keskivirhe			0,56	

**Taulukko 5.** MM-estimoinnin tulokset (Espoo).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Pinta-ala	-0,80	0,03	-23,40	0,00
Etäisyys Helsingin ydinkeskustaan	-1,14	0,11	-10,00	0,00
Asuntojen hinta-indeksi	2,48	0,07	36,60	0,00
Koko kiinteistö / määräala dummy	-0,26	0,05	-4,83	0,00
Selitysaste (keskistämätön malli)			-	
Regression keskivirhe			0,43	

## 3.2 *Empiirisen estimoinnin tulokset*

### 3.2.1 *Espoon hintamallit*

Taulukossa 3 on dokumentoitu pienimmän neliösumman mukaisen regressioanalyysin tulokset Espoon maanhinta-aineistolla, kun mallirakenteena on käytetty tavallista joustomallia, ja kun poikkeavia havaintoja ei ole kontrolloitu lainkaan. Estimoidun mallin elinvoimaisuus on yleisesti ottaen varsin heikko: hintamallin selitysaste on 58 % ja jäännöskeskijointa 0,71; molemmat ekonometriset perustunnusluvut selvästi huonompia kuin kiinteistöarvioinnissa yleisesti sovelletut raja-arvot. Hedonisen mallin jäännökset eivät ole normaalijakautuneita, vaan residuaalijakauma on selvästi huipukas sekä virhevarianssi saattaa olla heteroskedastinen. Lisäksi RESET-testi, jolla testataan mallin mahdollisia spesifiointivirheitä, osoittaa, että mallimuuttujien välisissä riippuvuussuhteissa on havaittavissa epälineaarisuutta, jota tavallinen joustomalli ei kykene täysin pätevästi ottamaan huomioon. Estimoidut hedoniset hinnat näyttäisivät olevan stabiileja ajan osalta, joista Helsinki-etäisyyden ja asuntojen hintaindeksin vaikutus näyttäisi olevan ylijoustavaa (eli ao. parametrien itseisarvot ovat suurempi kuin 1) ja pinta-alan sekä luokkamuuttujan vaikutus alijoustavaa (eli ao. parametrien itseisarvot ovat pienempiä kuin 1).

Joustomallissa kukin estimoitu hedoninen hinta antaa vastauksen kysymykseen, kuinka monta prosenttiyksikköä vaste muuttuu, jos selittäjän arvo muuttuu tietyn prosenttimäärän. Taulukoissa 3, 4, 6 ja 7 regression keskivirhe kuvaa tavallista jäännöskeskijointaa regressiopinnan ympärillä.

Taulukossa 4 on dokumentoitu pienimmän neliösumman mukaisen regression tulokset Espoon maanhinta-aineistolla, kun mallirakenteena on käytetty tavallista joustomallia, ja kun poikkeavat havaintopisteet ovat poistettu analyysissä kokonaan. Tässä poikkeavien havaintopisteiden kriteerinä on sovellettu standardoidun jäännöksen arvoa 2,5, joita suuremmat arvot osoittavat, että kyseinen piste on muukalainen; tämä raja-arvo 2,5 on usein sovellettu kiinteistömarkkinoilla (ks. esim. Thorson 1994). Poikkeavia havaintopisteitä voidaan identifioida kolme kappaletta ja estimoidut regressiokertoimet ja näiden tilastollinen merkittävyys muuttuvat hieman taulukon kolme hedoniseen malliin verrattaessa. Näin saadun mallin selitysaste nousee selvästi (suhteellinen parannus n. 17 % ja absoluuttinen parannus 10 %) ja regression keskivirhe laskee oleellisesti (suhteellinen parannus n. 21 % ja absoluuttinen parannus n. 15 %). Mallin jäännöstermien tarkastelu osoittaa, että niiden jakauma on vielä hieman huipukkaampi kuin normaalijakauman tapauksessa, ja että virhetermit olisivat heteroskedastisia. RESET-testi kertoo, että mallimuuttujien välisissä riippuvuussuhteissa on havaittavissa epälineaarisuutta, jota käytetty mallirakenne ei täysin pysty huomioon ottamaan. Lisäksi estimoidut hedoniset hinnat näyttäisivät olevan varsin epästabiileja. Helsinki-etäisyyden ja asuntojen hintaindeksin vaikutus näyttäisi olevan ylijoustavaa ja pinta-alan sekä luokkamuuttujan vaikutus alijoustavaa.

Taulukossa 5 on esitetty 3-vaiheisen MM-estimoidun hedonisen hintamallin tulokset Espoon maanhinta-aineistolla, kun mallirakenteena on käytetty tavallista joustomallia, ja kun poikkeavat havainnot ovat kontrolloitu painottamalla havaintopisteitä havaittujen virheiden mukaan. Poikkeavia havaintopisteitä ei siis ole poistettu MM-estimoinnissa, mutta riittävän poikkeavat ääriarvot saavat painon nolla, jolloin efektiivisesti joidenkin havaintojen vaikutus on sama, kuin jos nämä olisi poistettu hinta-aineistosta. Parametrien arvot ja näiden tilastollinen merkitsevyys muuttuvat jossain määrin; erityisen selvästi erottuvat nyt asuntojen hintaindeksin ja pinta-alan tilastolliset vaikutukset (t-arvot ovat korkeat: 36,6 ja -23,4, vastaavasti). Regressiomallin keskivirhe on suhteellisesti lähes 40 % pienempi ja absoluuttisesti 28 % pienempi kuin taulukon 3 maanhintamallissa ja suhteellisesti noin 23 % pienempi ja absoluuttisesti 13 % pienempi kuin taulukon 4 mukaisessa hintamallissa, ts. hedonisen mallin sisäinen tarkkuus näyttäisi kasvavan oleellisesti vikasietoisessa vaihtoehdossa. Selitysaste ei ole tavallisessa mielessä laskettavissa, kun MM-estimointia käytetään. Regression keskivirhe on taulukossa 5 ja 8 toisessa vaiheessa estimoitujen jäännösten hajonta. Helsinki-etäisyyden ja asuntojen hintaindeksin vaikutus näyttäisi edelleen olevan ylijoustavaa ja pinta-alan sekä luokkamuuttujan vaikutus alijoustavaa.

### 3.2.2 Nurmijärven hintamallit

Taulukossa 6 on dokumentoitu pienimmän neliösumman mukaisen regression tulokset Nurmijärven kunnan maanhinta-aineistolla, kun mallirakenteena on käytetty tavallista joustomallia, ja kun poikkeavia havaintoja ei ole kontrolloitu lainkaan. Estimoidun mallin elinvoimaisuus on yleisesti ottaen heikko: Hintamallin selitysprosentti on alhainen (44 %) ja jäännöskeskiahajonta on korkea (0,56). Hieman täsmällisemmin ilmaistuna molemmat tunnusluvut kertovat, että hintojen havaitusta vaihtelusta valtaosa olisi satunnaisvaihtelua. Virheiden jäännöstarkastelu osoittaa, että virheiden jakauma on huipukkaampi kuin normaalijakauma ja, että virheet olisivat heteroskedastisia. RESET-testi ei paljasta, että sellittämätöntä epälineaarisuutta olisi havaittavissa estimoidussa mallirakenteessa. Estimoidut hedoniset hinnat näyttäisivät lisäksi olevan kohtuullisen stabiileja ajan suhteen. Ainoastaan asuntojen hintaindeksin (ja vakiomuuttujan) vaikutus neliöhintoihin on ylijoustavaa ja muiden muuttujien vaikutus on puolestaan alijoustavaa.

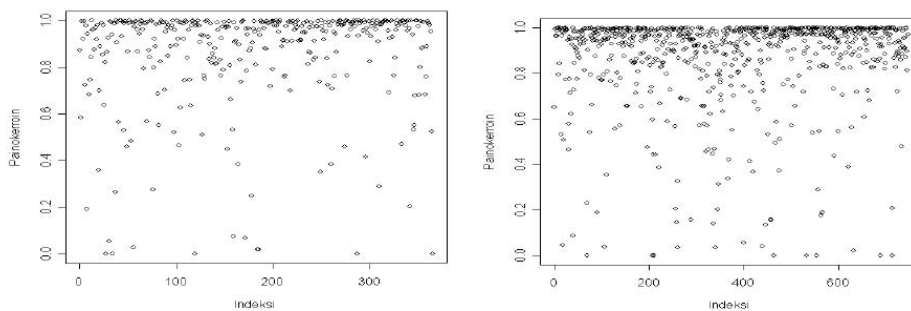
Taulukossa 7 on dokumentoitu pienimmän neliösumman mukaisen regression tulokset Nurmijärven kunnan maanhinta-aineistolla, kun mallirakenteena on käytetty tavallista joustomallia, ja kun poikkeavat havaintopisteet ovat poistettu analyysistä kokonaan käyttäen raja-arvona havaitun virheen standardoitua arvoa 2,5. Tällöin poikkeavia havaintopisteitä voidaan identifioida ainoastaan yksi kappale. (Se, että poikkeavia havaintoja identifioidaan ainoastaan yksi kappale Nurmijärven aineistolla saattaa johtua siitä, että poikkeavat havainnot estävät toisiaan näkymästä.) Estimoidut regressiokertoimet ja näiden tilastollinen merkitsevyys muuttuvat hieman taulukon 6 hedoniseen malliin verrattaessa. Sen sijaan estimoidun mallin selitysaste ja regression keskivirhe pysyvät suunnilleen

muuttumattomina. Jäännösten jakauma on hieman huipukkaampi kuin normaali-jakauma ja virheet näyttäisivät olevan myös heteroskedastisia. RESET-testi ei paljasta, että selittämätöntä epälineaarisuutta olisi havaittavissa estimoidussa mallirakenteessa. Estimoidut hedoniset hinnat näyttäivät lisäksi olevan kohtuullisen stabiileja ajan suhteen. Ainoastaan asuntojen hintaindeksin (ja vakiomuuttujan) vaikutus neliöhintoihin on ylijoustavaa ja muiden muuttujien vaikutus on puolestaan alijoustavaa.

Taulukossa 8 on esitetty kolmivaiheisen MM-estimoidun hedonisen hintamallin tulokset Nurmijärven kunnan maanhinta-aineistolla, kun mallirakenteena on sovellettu tavallista tulomuotoista joustomallia, ja kun poikkeavat havainnot ovat kontrolloitu painottamalla havaintopisteitä havaittujen virheiden mukaan. Estimoidut hedoniset hinnat muuttuvat verrattuna taulukoiden 6 ja 7 arvoihin. Muuttujien regressiokertoimien vaikutukset näyttäisivät kasvaneen (erityisesti pinta-alan ja vakiotermin vaikutus) lukuun ottamatta luokkamuuttujaa, jonka vaikutus on nyt hieman pienempi. Myös muuttujien tilastollisessa merkittävydessä on selviä eroja verrattaessa taulukoiden 6 ja 7 arvoihin. Regressiomallin keskivirhe on suhteellisesti lähes 25 % pienempi ja absoluuttisesti n. 14 % pienempi kuin taulukon 6 ja 7 mukaisissa hintamalleissa, ts. hedonisen mallin sisäistä tarkkuutta voidaan oleellisesti parantaa vikasietoisessa vaihtoehdossa. Ainoastaan asuntojen hintaindeksin (ja vakiomuuttujan) vaikutus neliöhintoihin on ylijoustavaa ja muiden muuttujien vaikutus on puolestaan alijoustavaa.

Kuvassa 1 on esitetty kunkin havainnon suhteellinen paino Espoon ja Nurmijärven kunnan maanhinta-aineistossa; nämä ovat MM-estimoinnin kullekin havainnolle antamat painot lukuvälin  $[0, 1]$  joukossa. Espoon hinta-aineistossa on kymmenen havaintopistettä, jotka efektiivisesti saavat painon nolla tai arvon, joka on hyvin lähellä nollaa. Sen sijaan Nurmijärven hinta-aineistolla kyseisiä havaintopisteitä on 16 kappaletta. Kuvasta ilmenee, että valtaosa havaintopisteistä molemmilla hinta-aineistolla saa arvon yksi eli normaalin pienimmän neliösumman mukaisen painon.

**Kuva 1.** Painot Espoon (vas.) ja Nurmijärven (oik.) havaintoaineistoilla.



**Taulukko 6.** Pienimmän neliösumman regression tulokset (Nurmijärvi; poikkeavat havainnot mukana).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Vakio	2,87	0,76	3,78	0,00
Pinta-ala	-0,63	0,03	-18,80	0,00
Etäisyys Helsingin ydinkeskustaan	-0,95	0,13	-7,22	0,00
Etäisyys Nurmijärven kirkonkylään	-0,24	0,04	-5,39	0,00
Asuntojen hinta-indeksi	1,51	0,11	13,90	0,00
Koko kiinteistö / määräala dummy	-0,20	0,04	-4,51	0,00
Selitysaste			0,44	
Regression keskivirhe			0,56	

**Taulukko 7.** Pienimmän neliösumman regression tulokset (Nurmijärvi; poikkeavat havainnot poistettu).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Vakio	2,74	0,77	3,55	0,00
Pinta-ala	-0,61	0,03	-18,10	0,00
Etäisyys Helsingin ydinkeskustaan	-0,93	0,13	-6,95	0,00
Etäisyys Nurmijärven kirkonkylään	-0,23	0,04	-5,15	0,00
Asuntojen hinta-indeksi	1,48	0,11	13,40	0,00
Koko kiinteistö / määräala dummy	-0,20	0,04	-4,54	0,00
Selitysaste (keskistämätön malli)			0,42	
Regression keskivirhe			0,57	

**Taulukko 8.** MM-estimoinnin tulokset (Nurmijärvi).

Muuttuja	Regressio-kerroin	Keskivirhe	t-arvo	p-arvo
Vakio	3,70	0,66	5,60	0,00
Pinta-ala	-0,77	0,03	-26,71	0,00
Etäisyys Helsingin ydinkeskustaan	-0,91	0,11	-7,95	0,00
Etäisyys Nurmijärven kirkonkylään	-0,23	0,04	-5,93	0,00
Asuntojen hinta-indeksi	1,57	0,09	16,63	0,00
Koko kiinteistö / määräala dummy	-0,14	0,04	-3,69	0,00
Selitysaste (keskistämätön malli)			-	
Regression keskivirhe			0,43	

#### 4 Ennustustarkkuuden tarkastelu

Ennustustarkkuus on tyypillisesti kriittinen tekijä hedonisen mallin empiirisen pätevyyden arvioinnissa käytännön sovelluksissa. Ennustustarkkuutta voidaan mitata lukuisilla erilaisilla indikaattoreilla (esim. Case ym. 2004), ja hedonisten mallien suhteellinen suorituskyky vaihtelee käytettyjen tunnuslukujen mukaan. Kuitenkaan kaikkia tunnuslukuja, jotka liittyvät ennustustarkkuuden mittaamiseen, ei voida pitää yhtä informatiivisina hedonisessa mallitarkastelussa: Ensimmäisen tärkeitä mittareita ovat ne kriteerit, jotka liittyvät keskivirheen tarkasteluun (eli *keskimääräisen* virheen arviointiin, joka mittaa mallin tulosten osuvuutta), ja kriteerit, jotka tuottavat informaatiota tilastollisen yhteyden voimakkuudesta ennusteiden ja todellisten arvojen välillä. Varianssipohjaiset indikaattorit – vaikkakin yleisesti käytettyjä – ovat usein toisarvoisia monissa käytännön maanarviointitilanteissa.

Keskimääräistä ennustevirhettä on arvioitu tässä osiossa keskimääräisellä aritmeettisella ennustevirheellä, mediaaniennustevirheellä, painotetulla keskimääräisellä virheellä, painotetulla mediaanivirheellä, virhekeskipisteellä, suhteellisella keskimääräisellä virheellä ja suhteellisella mediaanivirheellä. Sääntönä on, että mitä pienempiä arvoja *ao.* keskivirheen kriteerit tuottavat, sitä paremmasta tilastollisesta mallista on kysymys. Keskiarvopohjaisten osuvuusindikaattorien käyttö on usein maanhintatutkimuksissa varteenotettavampi vaihtoehto kuin mediaanipohjaisten keskilukujen soveltaminen, sillä jälkimmäiset mittarit painottavat liian voimakkaasti tyypillisiä virheitä suhteessa suurin tai pieniin virheisiin, koska nämä jättävät huomioon ottamatta relevanttia informaatiota virheiden välisistä etäisyyksistä. Suuri ongelma käytettäessä absoluuttisia keskilukuja on, että näiden tulokset riippuvat käytetystä mittakaavasta, jolloin vertailu erilaisten mallirakenteiden välillä saattaa muodostua ongelmaksi. Siksi kaksi suhteellista keskivirheen mittaria – suhteellinen keskimääräinen virhe ja suhteellinen mediaanivirhe – on dokumentoitu myös taulukoissa 9 ja 10.

Kolme eri mittaria tilastollisen yhteyden voimakkuudesta mallin ennusteiden ja todellisten arvojen välillä on raportoitu. Ensinnäkin on laskettu tavallinen Pearsonin korrelaatiokerroin, joka on hyödyllinen tilastollisen relaation mittari, kun virheet ovat normaalijakautuneita ja huomio on virheiden yhteisvaihtelulla. Mitä korkeampi Pearsonin korrelaatiokerroin on ennusteiden ja todellisten arvojen välillä, sitä parempi on tilastollinen malli. Merkittävä ongelma klassisen korrelaation käytölle maanhintatutkimuksissa on tämän voimakas riippuvuus normaalijakaumaoletuksesta, joka ei tyypillisesti pidä paikkaansa johtuen poikkeavista virhetermeistä, jotka neliöidään korrelaatiokertoimen nimittäjässä. Tämä puolestaan johtaa samansuuruisiin hajontalukuihin eri mallivaihtoehtojen välillä. Näin ollen klassisen korrelaatiokertoimen käyttö tuottaa, ei ainoastaan epäpäteviä tuloksia, mutta myös tuloksia, joita on vaikea asettaa paremmuusjärjestykseen. Skaalatun sisätulon käyttö on suositeltavampi vaihtoehto, joka mittaa ennusteen voimakkuutta todellisten hintojen suunnassa (sisätulolla mitattuna), ja joka on skaalattu keskistämättömien hintojen kokonaisuajonnalla. Mitä kor-

keampi skaalattu sisätulo on, sen parempi tilastollinen malli on kyseessä. Skaalattun sisätulon käytön on empiirisesti huomattu toimivan hyvin tilastollisen yhteyden voimakkuuden mittarina mallin ennusteiden ja todellisten arvojen välillä (ks. tarkemmin Hannonen 2006b). Se tuottaa osuvampia tuloksia, jotka eivät ole voimallisesti riippuvia tietyistä jakaumaoletuksista. Myös gravitaatio (McMillen 2001) on raportoitu, mikä saadaan jakamalla mallin ennusteiden ja todellisten havaintopisteiden painollinen sisätulo näiden euklidisella etäisyysmitalla. Mitä suuremman arvon gravitaatio saa, sitä parempi tilastollinen malli on kyseessä, koska tällöin yhteys ennusteiden ja todellisten arvojen välillä on voimakkaampi kuin alhaisemman gravitaation tapauksessa. Gravitaatio näyttää olevan elinvoimainen yhteyden voimakkuuden mittari, joka tuottaa tyypillisesti skaalatun sisätulon kanssa yhteneväisiä tuloksia.

Juurikeskineliövirhe (RMSE) on kenties tavallisimmin käytetty numeerisen ennustetarkkuuden indikaattori, joka pääsääntöisesti kontrolloi ennusteiden luotettavuutta (reliabiliteettia), ei tosiasiallista ennusteiden osuvuutta (validiteettia). Lisäksi tämä tunnusluku on äärimmäisen herkkä poikkeaville havainnoille. Kyseinen tunnusluku näyttää liioittelevan ennustusvirheiden varianssia tapauksissa, joissa esiintyy joitakin suuria ennustevirheitä (mikä on tyypillistä maanhintatutkimuksissa). Tämä herkkyys ääriarvoille saattaa myös aiheuttaa, että juurikeskineliövirhe tuottaa tuloksia, joita on vaikea asettaa paremmuusjärjestykseen. Keskimääräinen absoluuttinen virhe (MAE) on yleisesti tarkoituksenmukaisempi ennusteiden luotettavuusindikaattori, joka soveltuu poikkeavien ennustevirheiden käsittelyyn. Yksinkertainen ja hyödyllinen ennusteiden luotettavuusindikaattori on ennusteväli: Mitä lyhyempi väli on, sitä vähemmän vaihtelua ennusteissa on odotettavissa eli sitä luotettavampi tilastollinen malli on. Laajasti sovellettu ennustetarkkuuden mittari on suhteellinen keskimääräinen absoluuttinen virhe (MAPE), mikä on raportoitu yhdessä ”vikasietoisien” muunnelman kanssa. Näitä on myös kritisoitu epäsymmetrisyyden ja epästabiilisuusongelmista, kun havaintoaineisto saa pieniä arvoja. Mitä pienempiä arvoja juurikeskineliövirhe, keskimääräinen absoluuttinen virhe ja suhteellinen keskimääräinen absoluuttinen virhe sekä näiden mediaanivastineet (juurimediaanineliövirhe, mediaaninen absoluuttinen virhe ja suhteellinen mediaaninen absoluuttinen virhe) tuottavat, sitä paremmasta tilastollisesta mallista on kysymys.

Taulukossa 9 on esitetty kolmella eri perusteella estimoidun hedonisen mallin ennustetarkkuutta kuvaavia tunnuslukuja. PNS #1 viittaa pienimmän neliösumman keinolla estimoituun tavalliseen joustomalliin, missä poikkeavat havainnot ovat mukana estimoidussa mallissa, eikä näitä ole täten kontrolloitu millään erityisellä tavalla. PNS #2 kuvaa pienimmän neliösumman keinolla estimoitua tavallista joustomallia, missä poikkeavat havaintopisteet on poistettu mallista. Poikkeaviksi havaintopisteiksi on määritelty tapaukset, joiden standardoitu pienimmän neliösumman mukainen residuaali on suurempi kuin 2,5. Taulukon 9 tulokset kertovat selkeästi, että MM-estimointi parantaa mallin tulosten osumatarkkuutta huomattavasti, sillä kaikki seitsemän validiteettipohjaista indikaattoria (keskimääräinen aritmeettinen ennustevirhe, mediaaniennustevirhe,

painotettu keskimääräinen virhe, painotettu mediaanivirhe, virhekeskipiste, suhteellinen keskimääräinen virhe ja suhteellinen mediaanivirhe) ovat selvästi alhaisempia kuin muissa mallintamisvaihtoehdoissa.

**Taulukko 9.** Ennustustarkkuuden arviointi eri hintamalleilla – Espoon hintamalli.

Ennustustarkkuuden mittari	Hedoninen malli		
	PNS #1	PNS #2	MM-estimointi
Keskimääräinen aritmeettinen ennustevirhe	-0,36	-0,40	-0,13
Mediaaniennustevirhe	-0,14	-0,18	0,017
Painotettu keskimääräinen virhe	-2,98	-3,28	-1,06
Painotettu mediaanivirhe	-1,13	-1,48	0,16
Virhekeskipiste	-0,36	-0,40	-0,13
Suhteellinen keskimääräinen virhe	13,52	14,83	4,99
Suhteellinen mediaanivirhe	5,04	6,54	0,63
Ennusteväli [min, max]	[-2,44, 0,90]	[-2,46, 0,84]	[-2,61, 1,10]
Keskimääräinen absoluuttinen virhe	0,51	0,52	0,54
Mediaaninen absoluuttinen virhe	0,24	0,24	0,29
Suhteellinen keskimääräinen absoluuttinen virhe	65,74	66,71	69,83
Suhteellinen mediaaninen absoluuttinen virhe	8,74	9,42	10,23
Juurikeskineliövirhe	0,78	0,79	0,82
Juurimediaanineliövirhe	0,24	0,24	0,29
Korrelaatio (Pearson)	0,85	0,85	0,84
Skaalattu sisätulo	1,04	1,05	1,04
Gravitaatio	4 590,41	4 566,78	4 380,86

Erityisesti keskimääräinen aritmeettinen ennustevirhe, painotettu keskimääräinen virhe ja virhekeskipiste ovat jokainen n. 64 % alhaisempia kuin seuraavaksi parhaassa vaihtoehdossa (PNS #1) sekä suhteellinen keskimääräinen virhe n. 63 % alhaisempi kuin seuraavaksi parhaassa vaihtoehdossa (PNS #1). Mediaanipohjaisissa keskiluvuissa systemaattisen ennustevirheen pudotus on vieläkin selvempi MM-estimoinnin kohdalla. Sen sijaan varianssipohjaisissa indikaattoreissa (ennusteväli, keskimääräinen absoluuttinen virhe, mediaaninen absoluuttinen virhe, suhteellinen keskimääräinen absoluuttinen virhe, suhteellinen mediaaninen absoluuttinen virhe, juurikeskineliövirhe ja juurimediaanineliövirhe), jotka mittaavat ennusteiden luotettavuutta, on ainoastaan pieniä eroja eri lähestymistapojen välillä; pääsääntöisesti PNS #1 näyttäisi tuottavan hieman (muutaman prosenttiyksikön verran) luotettavampia tuloksia, sillä kaikki seitsemän reliabilitteittimittaria tuottavat hieman alhaisempia arvoja PNS #1:n kohdalla. Tilastollisen yhteyden voimakkuutta ennusteiden ja todellisten arvojen välillä kuvaavat tunnusluvut (korrelaatiokerroin, skaalattu sisätulo ja gravitaatio) sen sijaan kertovat, että optimaalisin lähestymistapa olisi joko PNS #1 tai PNS #2 MM-estimoinnin sijasta. Erityisesti gravitaatio on noin 5 % pienempi MM-estimoinnin kohdalla.

Taulukossa 10 on dokumentoitu PNS #1, PNS #2 ja MM-estimoinnin ennustustarkkuutta kuvaavat tunnusluvut Nurmijärven kunnan hinta-aineistolla. Nyt optimaalisen lähestymistavan valinta on hieman ongelmallisempi kuin taulukon 9 tapauksessa. Tarkempi kokonaisarviointi kuitenkin paljastaa, että MM-estimointi tuottaa nyt heikompia tuloksia kuin muut vaihtoehdot. Erityisesti mallin tulosten osuvuutta kuvaavissa virhekeskiluvuissa on selkeää heikennystä keskimääräisen aritmeettisen ennustevirheen, painotetun keskimääräisen virheen, virhekeskipisteen ja suhteellisen keskimääräisen virheen osalta. Esimerkiksi, kun verrataan PNS #1 ja MM-estimointia keskenään, niin ensimmäisessä lähestymistavassa keskimääräinen aritmeettinen ennustevirhe on n. 65 % pienempi, painotettu keskimääräinen virhe n. 60 % pienempi, virhekeskipiste n. 60 % pienempi ja suhteellinen keskimääräinen virhe n. 66 % pienempi. Sen sijaan mediaanipohjaisissa indikaattoreissa on varsin pieniä eroja PNS #2 eduksi. Kuitenkin mediaanipohjaisia tunnuslukuja voidaan pitää sekundäärisinä, sillä nämä hukkaavat arvokasta informaatiota etäisyyksistä eivätkä huomioi riittävässä määrin suuria ja pieniä virheitä, jotka ennusteaineistossa on luonnollisesti oletettava todellisiksi arvoiksi.

**Taulukko 10.** Ennustustarkkuuden arviointi eri hintamalleilla – Nurmijärven hintamalli.

Ennustustarkkuuden mittari	Hedoninen malli		
	PNS #1	PNS #2	MM-estimointi
Keskimääräinen aritmeettinen ennustevirhe	-0,052	-0,061	-0,15
Mediaaniennustevirhe	0,041	0,029	-0,041
Painotettu keskimääräinen virhe	-0,54	-0,61	-1,36
Painotettu mediaanivirhe	0,36	0,26	-0,35
Virhekeskipiste	-0,060	-0,070	-0,15
Suhteellinen keskimääräinen virhe	3,57	4,18	10,41
Suhteellinen mediaanivirhe	2,92	2,04	2,90
Ennusteväli [min, max]	[-2,30, 1,05]	[-2,31, 1,03]	[-2,36, 0,91]
Keskimääräinen absoluuttinen virhe	0,43	0,43	0,40
Mediaaninen absoluuttinen virhe	0,25	0,26	0,23
Suhteellinen keskimääräinen absoluuttinen virhe	382,65	393,90	416,86
Suhteellinen mediaaninen absoluuttinen virhe	18,45	17,87	13,89
Juurikeskineliövirhe	0,65	0,65	0,64
Juurimediaaninelövirhe	0,25	0,26	0,23
Korrelaatio (Pearson)	0,67	0,67	0,68
Skaalattu sisätulo	0,86	0,87	0,93
Gravitaatio	1 919,47	1 911,88	2 077,00

Mallien luotettavuutta kuvaavissa varianssipohjaisissa indikaattoreissa on havaittavissa ainoastaan pieniä eroja. Ennusteväli on hieman kapeampi MM-estimoinnin tapauksessa ja keskimääräinen absoluuttinen virhe on n. 7 % pienempi kuin PNS #1 ja PNS #2 tapauksissa. Myös yleisesti sovellettu juuri-

keskineliövirhe on 1,5 % alhaisempi MM-estimoinnin tapauksessa. Sen sijaan suhteellinen keskimääräinen absoluuttinen virhe on hieman korkeampi MM-estimaattorien kohdalla. Tilastollisen yhteyden voimakkuutta ennusteiden ja todellisten arvojen välillä kuvaavista tunnusluvuista oleellimmat eli skaalattu sisätulo ja gravitaatio ovat MM-estimoinnin tapauksessa korkeampia (skaalattu sisätulo n. 7,5 % ja gravitaatio n. 8,5 %).

Ennustetarkastelu siis paljasti varsin selvästi, että MM-estimointi tuotti Espoon aineistolla parhaan tuloksen, koska erityisesti keskimääräistä ennustevirhettä voitiin selvästi pienentää pienimmän neliösumman keinoin (tapaus 1: poikkeavat havaintopisteet mukana mallissa ja tapaus 2: poikkeavat havaintopisteet poistettu mallista) verrattuna. Sen sijaan Nurmijärven hinta-aineistolla ennustetarkastelussa parhaan tuloksen tuotti pienimmän neliösumman keino, jossa hintamallista ei oltu poistettu poikkeavia havaintopisteitä.

## 5 Johtopäätökset

Lineaariset pienimmän neliösumman mukaiset estimaatit hedonisille hinnoille saattavat käyttäytyä ei-optimaalisesti, kun virheet eivät ole normaalijakautuneita, ja erityisesti, kun virheiden jakauma on huipukas, mikä on tyypillinen tilanne maanhinta-aineistossa. Vikasietoisia mallintamismenetelmiä, jotka eivät ole herkkiä poikkeaville havaintopisteille, on lukuisia eri vaihtoehtoja, joille kaikille on kuitenkin yhteistä, että nämä antavat pienemmän painon epätyypillisille havainnoille kuin tavallinen pienimmän neliösumman mukainen keino. Tässä artikkelissa on tarkasteltu hyvin vikasietoista ja laskentaintensiivistä estimointimenetelmää, kolmivaiheista MM-estimointia, maanhintojen ekonometrisessä analyysissä. Tutkimustulokset ovat erityisiä tarkasteltujen markkinoiden ja maankäyttölajin osalta, mutta sovellettu tutkimusmenetelmä on yleinen ja sovellettavissa laajemminkin erilaisissa kiinteistöarvioinnin ja -markkinoiden analyysissä.

Empiirinen maanhintojen mallintaminen paljasti selvästi, että estimoitujen mallien sisäistä tarkkuutta regression keskivirheen mielessä voitiin parantaa (suhteellisesti noin 25–40 %) molemmilla hinta-aineistoilla, kun siirryttiin MM-estimointiin perinteisen pienimmän neliösumman sijasta. Empiirinen ennustetarkastelu paljasti selkeästi Espoon hinta-aineistoilla, että systemaattista ennustevirhettä voidaan merkittävästi pienentää, jos käytetään MM-estimointitekniikkaa tavallisen pienimmän neliösumman keinon sijasta. Samaa johtopäätelmää systemaattisen ennustevirheen pienentymisestä ei voida kuitenkaan tehdä Nurmijärven kunnan hinta-aineistolla. Se, että ennustetulokset ovat erilaiset Espoon ja Nurmijärven osamarkkinoiden välillä kertoo selvästi, että maan hinnan muodostuminen ko. osamarkkinoilla on toisistaan poikkeavaa ja spesifistä kullekin osamarkkina-alueelle. Näin ollen yhden alueen ennustetarkastelua koskevat tutkimustulokset eivät ole yleistettävissä toisella osamarkkina-alueelle. Hintatekijöistä pinta-alan ja asuntojen hintaindeksin tilastollinen vaikutus voimistui molemmilla aineistoilla, kun sovellettiin MM-estimointia.

**Lähdeluettelo**

Case, B., J. Clapp, R. Dubin and M. Rodriguez (2004). Modelling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *Journal of Real Estate Finance and Economics*, Vol. 29, No. 2, 167–191.

Fox, J. (1997). *Applied Regression Analysis, Linear Models and Related Methods*, Sage Publications.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Mathematical Statistics.

Hannonen, M. (2005). An Analysis of Land Prices: A Structural Time-Series Approach. *International Journal of Strategic Property Management*, 9, 145–172.

Hannonen, M. (2006a). An Analysis of Trends and Cycles Using Wavelet transforms. *International Journal of Strategic Portfolio Management*, Vol. 10, No. 1, 1–21.

Hannonen, M. (2006b). Forecastability of Land Prices: A Local Modelling Approach. *Nordic Journal of Surveying and Real Estate Research* (to appear).

Heinonen, T. (1993). Asuinpientalo- ja rivitalotonttien hinta ja hintakehitys Suomen kaupungeissa v. 1985–1991. *Maanmittaushallituksen julkaisu n:o 70*.

Hiltunen, A. (2003). Rakentamattomien pientalokiinteistöjen hinnanmuodostuksesta; Analyysi asuinpientalokiinteistöjen kauppahinnoista eräillä Suomen seutukunta-alueilla. *Kiinteistöopin ja talousoikeuden julkaisuja*, Espoo.

Huber, P. J. (2004). *Robust Statistics*. John Wiley & Sons.

Janssen, C., B. Soderberg and J. Zhou (2001). Robust Estimation of Hedonic Models of Price and Income for Investment Property. *Journal of Property Investment and Finance*, Vol. 19, No. 4, 342–360.

Kantola, J. (1982). Johdatus ekonometriin kiinteistöarvioinnin sovellutuksiin. *Otapaino*, Espoo.

Kantola, J. (1983). *Ekonometrinen kiinteistöarviointi Suomessa*. Otapaino, Espoo.

Leväinen, K. I. (1991). A Calculation Method for a Site Price Index. *The Association of Finnish Cities*.

McMillen, D. P. (1996). One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. *Journal of Urban Economics*, 40, 100–124.

McMillen, D. P. (2001). Nonparametric Employment Subcenter Identification. *Journal of Urban Economics*, 50, 448–473.

Marazzi, A. (1993). *Algorithms, Routines and S Functions for Robust Statistics*. Wadsworth & Brooks/Cole Publishing Company.

Rousseeuw, P.J. and V. J. Yohai (1984). Robust Regression by Means of S-estimators. In *Robust and Nonlinear Time Series* (eds. Franke J., W. Hardle and D. Martin). *Lecture Notes in Statistics* 26, 256–272, Springer-Verlag.

---

Thorson, J. A. (1994). The Use of Least Median of Squares in the Estimation of Land Value Functions. *Journal of Real Estate Finance and Economics*, 8, 183–190.

Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, Vol. 15, No. 20, 642–656.

Wallace, N. E. (1996). Hedonic-Based Price Indexes for Housing: Theory, Estimation and Index Construction. *Economic Review*, Vol. 3, 34–48.

Zaman, A., P. J. Rousseeuw and M. Orhan (2001). Econometric Applications of High-Breakdown Robust Regression Techniques. *Economic Letters* 71, 1–8.