

Polttoaineiden kuluttajahintojen tilastollinen mallintaminen
päivä- ja asematasolla

Joonas Isoketo

Tilastotieteen pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Vuosi 2014

Isoketo, Joonas: Polttoaineiden kuluttajahintojen tilastollinen mallintaminen päivä- ja asematasolla.

Tilastotieteen pro gradu -tutkielma, 44 sivua
Syyskuu 2014

Tiivistelmä

Tässä työssä tutkitaan, miten liikennepolttoaineiden, 95E, 98E ja diesel, vähittäishinnat ovat muuttuneet sekä alueellisesti että ajallisesti. Alueellisessa hinnoittelussa tutkitaan erityisesti, mitkä tekijät ovat yhteydessä hinnanmuuttamispäätökseen. Lisäksi selvitetään, miten paikallinen hintakilpailu ja aseman oma hintahistoria selittävät tulevia hintoja. Työssä kuvaillaan vielä, onko seutukuntien ja polttoaineketjujen välillä löydettävissä eroavaisuuksia hinnoittelussa. Polttoaineiden hintojen ajallisessa tarkastelussa keskitytään puolestaan tunnistamaan, mitkä tekijät ovat yhteydessä polttoaineiden hintoihin yleisesti.

Aineistona työssä on käytetty suomalaista verkkosivustoa www.polttoaine.net, jossa on kuluttajien ilmoittamia päivittäisiä eri polttoainelaatujen hintoja asemakohtaisesti. Tätä on rikastettu muilla aineistoilla, kuten verotiedoilla, etäisyystiedoilla, kuluttajahinnoilla ja raaka-öljyn maailmanmarkkinahinnoilla. Kaikki tutkielmassa käytettävä aineisto on avointa dataa, joka on siis verkossa vapaasti saatavilla. Tutkimuskysymyksiin haetaan vastauksia tilastollisilla malleilla. Hinnanmuutospäätöksen mallintamisessa käytetään logistista regressiota, aseman hintojen mallintamisessa käytetään lineaarista regressioanalyysia ja polttoaineiden hintakehitys mallinnetaan autokorrelaatiokorjatuilla regressiomalleilla. Tilastollisena työkaluna käytettiin SAS- ja R-ohjelmistoja.

Tuloksena on, että eniten paikallisessa hinnoittelussa vaikuttaa aseman oma hintahistoria. Kilpailun vaikutus jää kohtalaiseksi. Yleinen hintakehitys puolestaan on eniten yhteydessä raakaöljyn maailmanmarkkinahintaan ja inflaatiosta johtuvaan aikatrendiin. Suuria seutukuntakohtaisia hintaeroja ei esiinny. Eri polttoainelaatujen hinnoittelu on melko samanlaista, bensiini 98E tosin on kilpailulle joustamattomin. Tutkimuskysymysten malleja voidaan käyttää ennustamiseen. Mallien selitysasteet ja diagnostiikkatarkastelut ovat kohtalaisen hyviä. Analyyseja voitaisiin vielä tarkentaa ja parantaa, jos käytössä olisi tiedot asemien ja jalostajien varastotasoista ja katetavoitteista.

Avainsanat: Polttoaineen hinta, Lineaarinen malli, Regressioanalyysi, Logistinen regressio, Pienimmän neliösumman menetelmä, Yleistetty pienimmän neliösumman menetelmä, Avoin data.

Sisältö

1	Johdanto	1
2	Aineisto	3
3	Regressioanalyysi	5
3.1	Usean selittäjän regressiomalli	5
3.2	Estimointi	6
3.3	Parametrien otantajakumat ja testit	8
3.4	Mallidiagnostiikka	9
3.5	Logistinen regressio	9
3.6	Yleistetty pienimmän neliösumman menetelmä	11
4	Tilastollinen mallintaminen ja tulokset	13
4.1	Tutkimuskysymys 1: Miten ennustaa tietyn aseman seuraavan päivän hintaa?	14
4.1.1	Hinnanmuuttamispäätöksen mallinnus	14
4.1.2	Uuden hinnan mallinnus	18
4.2	Tutkimuskysymys 2: Miten mallintaa polttoaineiden hinnankehitystä?	28
4.2.1	Lähtökohdat	28
4.2.2	Mallinnus	28
4.2.3	Tulosten tulkinta	31
4.3	Tutkimuskysymys 3: Millaisia ovat alueelliset erot polttoaineiden hinnoissa?	35
5	Johtopäätökset	41
	Lähteet	43

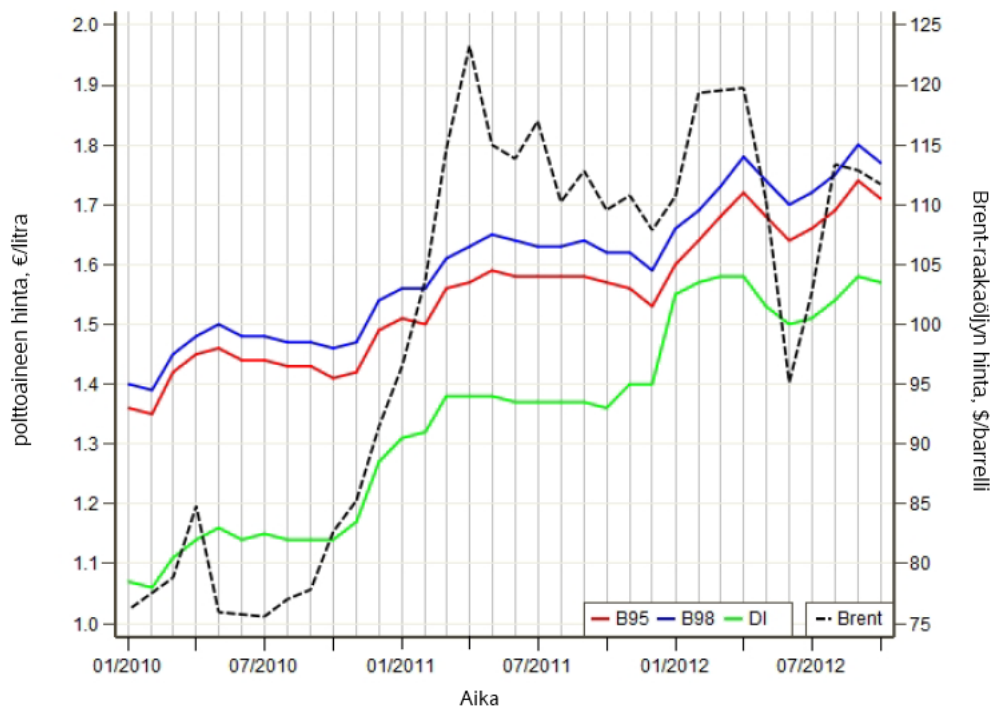
1 Johdanto

Pro gradu -tutkielmassani tarkastellaan lineaaristen regressiomallien avulla bensiini-laatuja 95E ja 98E sekä dieselin kuluttajahintaan vaikuttavia spatiaalisia eli alueellisia ja temporaalisia eli ajasta johtuvia tekijöitä. Tutkielman tavoitteena on rakentaa mahdollisimman hyviä tilastollisia malleja kuvaamaan polttoaineiden hinnoittelussa esiintyviä ilmiöitä. Analysoinnissa on käytetty avointa asema- ja päiväkohtaista hintadataa, jota on jalostettu rikkaammaksi niin ikään myös avoimista lähteistä.

Viimeiset vuodet polttoaineen korkea hinta on ollut suuri puheenaihe. Tämä ei tosin ole mikään ihme: öljytuotteet ovat yksi talouden tärkeimmistä hyödykkeistä. Hinnat ovat nyt huipussaan ja raakaöljyn maailmanmarkkinahinta on rikkonut ennätyksensä viimeisten vuosien aikana (Kuva 1). Liikenteen volyyymi on kasvanut suureksi, kun ihmiset liikkuvat yhä enemmän työmatkoillaan ja vapaa-ajallaan. Talous on voimakkaasti globalisoitunut, ja kotimarkkinoilla logistiikkaketju on pitkä ja materiaalivirrat suuria [15]. Tieliikenteessä tarvittavien polttomoottoreiden ylivoimaisesti tärkeimmät energialähteet ovat bensiini sekä diesel, joille ei toistaiseksi löydy haastajia polttoainemarkkinoilla. Kuluttajien kulutusrakenteessa liikkumiseen käytettävien polttoaineiden suora osuus on yksi suurimmista: auton tankkaaminen vie ison summan kuluttajien tuloista [14]. Lisäksi epäsuorat vaikutukset ovat huomattavia: hyödykkeiden hinnat sisältävät kuljetusten kustannukset. Öljyn hinnannousun vuoksi riippuvuutta siitä pyritään jatkuvasti vähentämään. Taloushistoriassa tunnetaan monta öljykriisiä, joilla on ollut suuria globaaleja vaikutuksia. Syinä näihin ovat olleet häiriöt raakaöljyn tarjonnassa [18].

Öljypohjaisten polttoaineiden lopulliseen hintaan vaikuttaa runsas määrä tekijöitä ja loppuhinnan määräytyminen on monimutkainen prosessi. Jos tarkastellaan bensiinin tapausta, tuotantoprosessi lähtee liikkeelle raakaöljyn tuotannosta. Ensimmäinen huomio tulee tarjontapuolelta: raakaöljyvarannot ovat rajoitetut ja niitä on vain tietyillä mailla. Öljyntuotanto on siis oligopolistista, ja sen merkittävin toimija on öljynviejämaiden yhteisorganisaatio OPEC. OPEC on kartelli ja pystyy näin ollen vaikuttamaan hintoihin kuluttajien kannalta epäsuotuisasti hintasopimuksilla ja tuotannon volyymin säätelyllä [18]. Vuonna 2011 suurimmat öljyntuottajat olivat Saudi-Arabia, Venäjä ja Yhdysvallat [20]. Raakaöljyä käytetään ympäri maailmaa, ja kysyntä on kasvanut ajan mittaan. Tulevaisuudessa kysynnän oletetaan kasvavan entisestään, kun kehittyvien maiden talouksien autokannat tulevat kasvamaan. Hintaan vaikuttavia tekijöitä kysynnän ja tarjonnan lisäksi ovat muun muassa kansainväliset suhteet, ympäristöpolitiikka, sääilmiöt, tuotantoteknologia, varastotasot, kysynnän vuodenaajoista johtuva kausivaihtelu, talouden suhdanteet ja tulevaisuuden ennusteet. Koska öljyä viedään tuottajamaista tuontimaihin, valuuttakurssien vaikutus näkyy hinnoissa. Raakaöljystä jalostetaan tuotteita käyttötarkoituksen mukaan, joten loppuhinta sisältää myös jalostajien marginaalit. Koska fyysiset matkat tuotantopaikalta loppumarkkinapaikalle ovat pitkiä, niin logistiikkakustannukset ovat myös osa loppuhintaa. Prosessin loppuvaiheessa hintaan vaikuttavat vielä verotus, maantieteellinen sijainti ja vähittäismyyntin paikallinen kilpailu. [20]

**POLTTOAINELAATUJEN 95E, 98E JA DIESEL SEKÄ BRENT-RAAKAÖLJYN
KUUKAUSITTAINEN KESKIHINTOJA VUOSINA 2010-2012**



Kuva 1: Polttoaineiden keskihintoja (vasen asteikko) ja Brent-raakaöljyn maailmanmarkkinahinta (oikea asteikko) vuosina 2010-2012. Lähteet: Tilastokeskus [14], U.S. Energy Information [17].

Vuoden 2010 alussa bensiinilaadut maksoivat keskimäärin noin 1.40 €/l, diesel puolestaan maksoi suunnilleen 1.10 €/l. Bensiinilaatujen litrahinnat pysyivät vuoden 2010 loppupuolella vielä alle 1.50 eurossa, diesel alle 1.20:ssä. Samaan aikaan Brent-raakaöljyn barrelihinta oli 80 dollarin tuntumassa. Arvonlisäverokantaa korotettiin 1.7.2011 yhdellä prosenttiyksiköllä. Samoihin aikoihin raakaöljyn hinta lähti huimaan nousuun: huhtikuussa 2011 raakaöljyn hinta kävi jo yli 120 dollarissa. Vuoden 2011 alussa 95E:n etanolipitoisuutta nostettiin viidestä prosenttiyksiköstä kymmeneen prosenttiyksiköön. Tästä alkoi myös polttoaineiden hinnannousu. Bensiinit maksoivat Q2/2011 jo 1.55 – 1.65 euroa litralta, diesel puolestaan lähes 1.40 euroa/l. Tämän jälkeen hinnoissa alkoi "sahausliike": kaikki edellä mainitut laskivat, nousivat, taas laskivat ja jälleen nousivat. Raakaöljy kävi tuona aikana vielä alle sadassa dollarissa, jolloin polttoaineidenkin hinnat notkahtivat hetkellisesti: 95E maksoi kesällä 2012 1.65 €/l, 98E 1.70 €/l ja diesel 1.50 €/l. Polttoaineiden valmisteveroja korotettiin vuoden 2012 alussa. Dieseliä korotettiin rankemmin kuin bensiinejä (katso Taulukko 1, s. 5). Tarkastelujakso päättyi näitä noin viisi senttiä korkeampiin hintoihin. Hintamuutokset tarkastelujakson alusta loppuun asti olivat suuria: 95E + 21 %, 98E + 21 %, diesel + 45 % ja raakaöljy + 43 % (USD).

Tutkimusongelmana ovat seuraavat kysymykset alikohtineen:

- Tutkimuskysymys 1: Miten ennustaa tietyn aseman seuraavan päivän hintaa?
 - Milloin asema muuttaa hintojaan?
 - Millainen yhteys on lähiasemien hinnoilla ja aseman edellisellä hinnalla aseman seuraavaan hintaan?
- Tutkimuskysymys 2: Miten mallintaa polttoaineiden hinnankehitystä?
 - Millainen yhteys on raakaöljyn maailmanmarkkinahinnalla, viikonpäivällä, vuodenaajalla ja verotuksen muuttumisella polttoaineiden keskihintaan?
 - Ovatko raakaöljyn hinnanmuutokset ja polttoaineen hinnanmuutokset symmetrisiä vai hyötyvätkö huoltoasemat laskevasta öljynhinnasta?
 - Onko dieselin talvilaatu kalliimpaa kuin kesälaatu? Nostiko siirtyminen 95E5-laadusta 95E10-laatuun 95E:n ja samalla 98E:n ja dieselin hintoja?
- Tutkimuskysymys 3: Millaisia ovat alueelliset ja ketjukohtaiset erot polttoaineiden hinnoissa?
 - Onko joissakin seutukunnissa halvempaa kuin muualla?
 - Alentaako paikallinen kilpailu hintasoa vai ilmeneekö viitteitä kartelleista?
 - Miten ketjujen hinnat eroavat toisistaan?

Näihin kysymyksiin pyritään pro gradu -tutkielmassa vastaamaan tilastollisilla menetelmillä. Tämän tutkielman luvussa 2 esitellään aineisto ja sen muuttujat. Luvussa 3 esitellään tutkimusongelmien ratkaisemiseen tarvittavaa teoriaa ja luvussa 4 esitellään osa tutkimuskysymyksiin vastaamiseen tarvittavista malleista ja niiden antamia empiirisiä tuloksia. Luku 5 käsittää johtopäätökset ja yhteenvedon.

2 Aineisto

Aineisto sisältää eri polttoainelaatujen asemakohtaisia päivähintoja. Tämän lisäksi aineistoon on yhdistetty eri kanavien kautta muuta tutkielman teossa tarvittavaa aineistoa kuten asemien maantieteelliset koordinaatit, valmiste- ja arvonlisäverotiedot, Brent-raakaöljyalaadun maailmanmarkkinahinta sekä valuuttakurssit. Edellä mainittujen avulla voidaan luoda uusia tarvittavia muuttujia, jotka esitellään myöhemmin. Muuttujien merkinnät esitellään myös myöhemmin kappaleessa 4. Tarkastelujakso on rajattu aikavälille 1.1.2010 – 2.10.2012 ja tarkasteluasemat on rajattu Helsingin, Jyväskylän, Kuopion, Lahden, Oulun, Tampereen ja Turun seutukuntiin (seutukuntaluokittelu [16, s. 50]). Syy rajaukseen on asemien osoitteiden validoinnin työläys. Analyysissä käytetyt Suomen keskihinnat kuitenkin pohjautuvat koko Suomen hintatietoihin.

Aineiston hintatiedot ovat peräisin Internet-sivuston <http://www.polttoaine.net> ylläpidon tietokannasta [11]. Sivustolta löytyy kootusti päivittäisiä eri polttoainelaatujen hintoja asemakohtaisesti. Hinnat ovat pääasiassa kuluttajien ilmoittamia. Sivusto on riippumaton polttoaineiden hintaseurantasivu, joka auttaa autoilijoita löytämään edullisimman bensiini- ja dieseljakeluaseman. Hinnat sisältävät verot eli arvonlisäveron ja valmisteveron. Verojen suuruudet tarkastelujaksolla on listattuna Taulukossa 1 [1], [10, 15§].

Aineistossa on 291 593 hintahavaintoa polttoainelaadusta 95E, 247 744 havaintoa polttoainelaadusta 98E ja 271 849 havaintoa dieselistä. Eri asemia aineistossa on 723 kappaletta. Asemien hintatietoja pidetään sivustolla näkyvillä viisi päivää ilman päivityksiä. Tämän vuoksi raaka-aineistoa on ensin laajennettu, koska aineistossa oli vain tietueet päivitysajankohdissa. Polttoaineiden vähittäismyyntiketjuja on yhdistelty, ja tarkastelussa on seitsemän kappaletta: ABC, Neste, SEO, Shell, ST1, Teboil ja Muu. Kunkin ketjun paitsi ketjun *Muu* nimen alla on myös pienempiä omalla nimellä toimivia ketjuja, jotka ovat sulautuneet edellä mainittuun ketjuun 2.10.2012 mennessä [19]. Ketju *Muu* sisältää sellaiset itsenäiset huoltoasemaketjut, jotka eivät ole sulautuneet edellä mainittuihin. Esimerkkinä mainittakoon, että ketju *Neste* sisältää A24-asemat ja ketju *Muu* sisältää muun muassa Ritoil-asemat.

Tutkielmassa raakaöljyn maailmanmarkkinahinnan aikasarjan tiedot on ladattu Yhdysvaltojen ympäristöministeriön alaisen laitoksen *U.S Energy Information Administrationin* (EIA) Internet-sivustolta [17]. Hinnat ovat niin ikään päivittäisiä ja ne ovat keskiarvoja kyseisen päivän spot-hinnoista. Hinnat ovat FOB-hintoja ('free on board') eli niihin sisältyy kuljetus- ja jakelupalvelut kyseisen tuontimaan rajalle saakka sekä mahdolliset vientitavaroiden verot vähennettynä tukipalkkioilla. Raakaöljyalaadusta on valittu Luoteis-Euroopassa jalostettu Brent-laatu, jota suurin osa Suomeen tuodusta öljystä on. Hintojen tilastointi ja julkaisu on yleistä, ja päivän hinnan kuluttajatkin näkevät helposti esimerkiksi sanomalehtien talousuutisista. Tämän raakaöljyalaadun hintasarja on ilmoitettu Yhdysvaltojen dollareissa. Kyseiset hinnat on muunnettu euroiksi käyttäen Suomen Pankin julkaisemia päivittäisiä valuuttakursseja [13]. Koska raaka-ainepörssi on viikonloput suljettu, niin lauantain ja sunnuntain kohdalla käytetään kyseisen viikon perjantain hintaa. Näin tehdään myös viikonloppun valuuttakursseille. Raakaöljyn maailmanmarkkinahinnassa näkyvät globaalit öljymarkkinailmiöt, jolloin tilastomenetelmillä oletettavasti päästään tehokkaammin käsiksi Suomen kotimarkkinoilla tapahtuviin arvonlisäyksiin.

Internet-sivuston [poltttoaine.net](http://www.polttoaine.net) hintatiedoista löytyy aseman katuosoite. Tämän perusteella asemille on liitetty maantieteelliset koordinaatit geokoodaamalla käyttäen Internet-sovellusta <http://www.findlatitudeandlongitude.com> [6]. Pituus- ja leveyspiirikoordinaattien avulla asemille laskettiin *R*-paketin '*fields*' funktiolla '*rdist.earth()*' [8] etäisyysmatriisi, jonka avulla voidaan tarkastella asemien lähialueen hintoja ja lukumääriä. Lähes kaikki muu aineiston käsittely on tehty SAS[®]-ohjelmiston versiolla 9.3.

Taulukko 1: Polttoaineisiin sovelletut verot 1.1.2010 – 2.10.2012

Aika	ALV, %	Valmistevero	
		Bensiini, €/l	Diesel, €/l
1.1.2010 - 30.6.2010	22	0.627	0.364
1.7.2010 - 31.12.2011	23	0.627	0.364
1.1.2012 - 2.10.2012	23	0.6504	0.4695

3 Regressioanalyysi

Johdannossa esitettyihin tutkimusongelmiin etsitään vastauksia regressioanalyysin avulla. Hintamuutospäätöksen mallinnuksessa käytetään logistista regressiota, uuden hinnan määrittämisessä lineaarista regressiota ja polttoaineiden hintakehitykseen valitaan autokorrelaatiokorjattu regressio. Niin ikään lineaarista regressiota käytetään, kun tutkitaan hyötyvätkö polttoaineasemat laskevasta raakaöljyn hinnasta. Tutkimuskysymys 3 analysoidaan kuvailevalla tilastotieteellä.

Regressioanalyysillä voidaan tutkia, miten tietty satunnaismuuttuja Y riippuu muuttujista X_1, \dots, X_p . Muuttujaa Y sanotaan selitettäväksi muuttujaksi tai vastemuuttujaksi. Muuttujia X_1, \dots, X_p sanotaan selittäviksi muuttujiksi tai regressoreiksi. Tässä työssä käsitellään regressiomalleja, joissa Y voi olla joko jatkuva tai binäärinen (dikotominen) muuttuja. Selittävät muuttujat voivat olla joko jatkuvia tai diskreettejä. Aliluvut 3.1–3.5 pohjautuvat kirjoihin [7] sekä [9] ja aliluku 3.6 kirjan [7] lisäksi artikkeliin [4].

3.1 Usean selittäjän regressiomalli

Oletetaan, että on mitattu n kappaletta havaintoja kustakin muuttujasta Y, X_1, \dots, X_p . Tällöin voidaan muodostaa Y -muuttujasta vektori $\mathbf{Y} = Y_i$, jossa i . alkio on mitatun muuttujan arvo havaintoyksiköllä i . Tehdään sama X -muuttujille. Edelleen vektoreista X_1, \dots, X_p voidaan rakentaa matriisi laittamalla jokainen vektori tämän matriisin sarakkeeksi. Kun ensimmäiseksi sarakkeeksi lisätään vielä ykkösiä sisältävä vektori $\mathbf{1}_{n \times 1}$, saadaan matriisi

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}.$$

Regressiomallissa muuttujajoukon X_1, \dots, X_p yhteys vastemuuttujan Y ilmaistaan havaintokohtaisesti yhtälöllä

$$(3.1) \quad Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \epsilon_i, \quad i = 1, \dots, n$$

ja yhtäpitävästi matriisimuodossa

$$(3.2) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{missä } \boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \dots \ \beta_p)^T \text{ ja } \boldsymbol{\epsilon} = (\epsilon_1 \ \dots \ \epsilon_n)^T.$$

Jokainen selittävä muuttuja X_1, \dots, X_p vaikuttaa vastemuuttujaan Y regressiokertoimensa β_j ($j = 1, \dots, p$) kautta. Lisäksi vastemuuttujalla Y on jokin X -muuttujista riippumaton taso β_0 . Jokaisella vektorin \mathbf{Y} alkiolla Y_1, \dots, Y_n on oma virhetermiansä ϵ_i . Virhetermi kuvaa sitä satunnaista osaa vastemuuttujasta Y , jota selittävät X_1, \dots, X_p muuttujat eivät pysty selittämään.

Virhetermeistä $\boldsymbol{\epsilon}$ oletetaan, että ϵ_i on riippumaton selittäjistä X_{1i}, \dots, X_{pi} kaikilla $i = 1, \dots, n$ ja että $E[\boldsymbol{\epsilon}] = \mathbf{0}$. Vielä lisäksi oletetaan, että virhetermien jakauma on tyypiltään normaalin ja varianssi on vakio kaikilla i . Tällöin siis $\epsilon_i \sim N(0, \sigma^2)$ ja $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$ kaikilla $i = 1, \dots, n$. Virhetermit ovat siis myös toisistaan riippumattomia. Vähentämällä vasteesta selittävien muuttujien vaikutus saadaan virhetermi:

$$Y_i - E[Y_i | X_1 = x_1, \dots, X_p = x_p] = \epsilon_i.$$

Kertoimet $\boldsymbol{\beta}$ ilmaisevat, miten selittävien X -muuttujien arvot vaikuttavat vastemuuttujaan Y . Kertoimille $\boldsymbol{\beta}$ saadaan tulkinta tarkastelemalla muuttujan Y ehdollisten odotusarvojen erotuksia. Nimittäin

$$\beta_j = E[Y | X_1 = x_1, \dots, X_j = x_j + 1, \dots, X_p = x_p] - E[Y | X_1 = x_1, \dots, X_j = x_j, \dots, X_p = x_p].$$

Tämän perusteella siis kerroin β_j ilmaisee, kuinka paljon keskimäärin Y muuttuu, kun selittävä muuttuja X_j kasvaa yhdellä yksiköllä ja muut selittävät muuttujat pysyvät ennallaan. Kerroin β_0 kuvaa, miten suuren arvon Y keskimäärin saa, kun kaikki X -muuttujat saavat arvokseen nolla. Edelliset laskut perustuvat odotusarvon lineaarisuuteen sekä oletukseen virhetermien odotusarvon asettamisesta nolaksi.

Nyt kun virhetermeiltä oletetaan edellä mainitut vaatimukset, voidaan määrittää muuttujan Y ehdollinen todennäköisyysjakuma ehdolla havaitut selittävät muuttujat X_1, \dots, X_p . Tämä jakauma on

$$Y | X_1 = x_1, \dots, X_p = x_p \sim N(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \sigma^2).$$

Ehdollinen tiheysfunktio tälle on

$$(3.3) \quad f(y_i | x_{i1}, \dots, x_{ip}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2}{2\sigma^2}}.$$

3.2 Estimointi

Edellä esiteltiin mallintamisen kannalta kiinnostavat $\boldsymbol{\beta}$ -kertoimet. Lisäksi esiintyi tuntematon virhetermien varianssi σ^2 . Kun tavoitteena on rakentaa tilastollinen malli kuvaamaan vasteen ja selittävien muuttujien riippuvuutta, niin tuntemattomat populaatioparametrit on estimoitava kerätyn aineiston avulla. Estimointi tehdään suurimman uskottavuuden menetelmällä. Määritetään siis parametrien $\boldsymbol{\beta}$ ja σ^2 uskottavuusfunktio ja etsitään sen maksimi. Lineaaristen mallien tapauksessa tämä johtaa pienimmän neliösumman menetelmään.

Esityksen (3.3) avulla saadaan \mathbf{Y} :n yhteisjakauma jokaisen vasteen Y_i jakaumien tulona. Näin voidaan tehdä, koska virheet ϵ_i oletettiin riippumattomiksi, jolloin myös vasteen alkiot y_i ovat ehdollisesti riippumattomia. Laskemalla tulo auki ja pitämällä yhteistiheysfunktioita parametrien $\boldsymbol{\beta}$ ja σ^2 funktiona saadaan uskottavuusfunktio

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 \right].$$

Kun tästä otetaan luonnollinen logaritmi, saadaan logaritminen uskottavuus

$$(3.4) \quad \ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2.$$

Uskottavuusfunktioilla $L(\boldsymbol{\beta}, \sigma^2)$ ja logaritmisella uskottavuusfunktioilla $\ell(\boldsymbol{\beta}, \sigma^2)$ on samat maksimipisteet $\boldsymbol{\beta}$ ja σ^2 . Maksimoidaan näistä $\ell(\boldsymbol{\beta}, \sigma^2)$. Etsitään aluksi estimaattorit regressiokertoimille $\boldsymbol{\beta}$. Toisin sanoen etsitään jokainen β_j siten, että $\ell(\boldsymbol{\beta}, \sigma^2)$ saa suurimman arvonsa. Yhtälöstä (3.4) huomataan, että funktio saa suurimman arvonsa, kun $\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$ saavuttaa pienimmän arvonsa. Otetaan käyttöön merkintä $S(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$, joka kuvaa virhetermien neliösummaa, joka tullaan minimoimaan. Tästä tulee estimointitavan nimitys pienimmän neliösumman menetelmä. Funktion $S(\boldsymbol{\beta})$ ääriarvot ovat osittaisderivaattojen nollakohdissa. Osittaisderivoidaan $S(\boldsymbol{\beta})$ jokaisen regressiokertoimen β_j suhteen, jolloin uskottavuusyhtälöistä saadaan $\boldsymbol{\beta}$:n suurimman uskottavuuden estimaattoriksi

$$(3.5) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

Varianssiparametri σ^2 voidaan estimoida yhtälöstä (3.4), kun $\boldsymbol{\beta}$ korvataan sen edellä lasketulla estimaatilla $\hat{\boldsymbol{\beta}}$. Tällöin logaritminen uskottavuusyhtälö tulee muotoon

$$\ell(\hat{\boldsymbol{\beta}}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{S(\hat{\boldsymbol{\beta}})}{2\sigma^2}.$$

Tästä osittaisderivoinnin kautta σ^2 :n suurimman uskottavuuden estimaattoriksi saadaan

$$(3.6) \quad \hat{\sigma}^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n} = \frac{1}{n} \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_p x_{ip} \right)^2.$$

3.3 Parametrien otantajakaumat ja testit

Jotta mallin antamista tuloksista voitaisiin tehdä johtopäätöksiä tarkemmin, on tunnettava regressiokertoimien estimaattoreiden otantajakaumat. Käytännön kannalta on kiinnostava tietää, eroaako jokin β_j tilastollisesti jostakin ennalta määritetystä vakiosta. Tätä päästään tutkimaan joko luottamusvälien tarkastelulla tai tilastollisella testillä.

Edellä on johdettu β :n pienimmän neliösumman estimaattori (3.5). Määritetään tämän otantajakauma. Aloitetaan laskemalla ensin $\hat{\beta}$:n odotusarvovektori:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon.$$

Koska oletettiin, että $E[\epsilon] = \mathbf{0}$ ja että muuttujat X_1, \dots, X_p eivät ole stokastisia, niin $E[\hat{\beta}] = \beta$. Siis $\hat{\beta}$ on β :n harhaton estimaattori.

Lähtökohtaisesti on oletettu virhetermien ϵ riippumattomuus ja normaalisuus. Riippumattomuudesta seuraa, että virhetermien kovarianssimatriisi $\text{Cov}[\epsilon] = \sigma^2 \mathbf{I}$. Lasketaan $\hat{\beta}$:n kovarianssimatriisi:

$$\begin{aligned} \text{Cov}[\hat{\beta}] &= E[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])^T] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon \epsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\epsilon \epsilon^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\sigma^2 \mathbf{I}] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned}$$

Koska vektorin ϵ oletetaan noudattavan n -ulotteista normaalijakaumaa $N_n(\mathbf{0}, \sigma^2 \mathbf{I})$, niin myös estimaattorin $\hat{\beta}$ otantajakauma on tällöin tyypiltään normaalin. Edellä laskettiin $\hat{\beta}$:n odotusarvovektori ja kovarianssimatriisi. Näin ollen

$$\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}).$$

Otetaan käyttöön merkintä $c_{ij} = [(\mathbf{X}^T \mathbf{X})^{-1}]_{ij}$ ($i, j = 1, \dots, p+1$). Tällöin $\hat{\beta}$:n j . alkionle pätee

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj}) \quad \Leftrightarrow \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 c_{jj}}} \sim N(0, 1).$$

Koska σ^2 on tuntematon, korvataan se estimaatillaan. Tällöin saadaan, että

$$(3.7) \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \sim t(n - p - 1).$$

Nyt kun $\widehat{\boldsymbol{\beta}}$:n otantajakauma tunnetaan, niin voidaan määrittää kaksisuuntainen testi, jolla testataan hypoteesia $\beta_j = b$, missä b on vakio. Asetetaan nolla- ja vastahypoteesi:

$$H_0: \beta_j = b, \quad H_1: \beta_j \neq b.$$

Jos H_0 pätee, niin

$$(3.8) \quad t = \frac{\widehat{\beta}_j - b}{\text{s.e}(\widehat{\beta}_j)} = \frac{\widehat{\beta}_j - b}{\sqrt{\widehat{\sigma}^2 c_{jj}}} \sim t(n - p - 1).$$

Olkoon testin merkitsevyytaso α . Testin mukaan H_0 hylätään, jos

$$|t_{\text{havaintu}}| > t_{\alpha/2, n-p-1}.$$

3.4 Mallidiagnostiikka

Edellä aliluvussa 3.1 tehtiin useita oletuksia mallia koskien. Jos mallia koskevat oletukset eivät ole voimassa, estimointitulokset voivat olla harhaisia, epätarkentuvia ja tehottomia. Kootaan yhteen regressioanalyysin perusoletukset:

1. *Lineaarinen funktionaalinen muoto: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$;*
2. *Jäännösten odotusarvovektori on nollavektori, jäännökset homoskedastisia, autokorreloimattomia ja normaalisti jakautuneita: $\boldsymbol{\epsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$;*
3. *Jäännösten ehdollinen odotusarvovektori nollavektori: $E[\boldsymbol{\epsilon} | \mathbf{X}] = \mathbf{0}$;*
4. *Selittävät muuttujat ovat lineaarisesti riippumattomia: \mathbf{X} on $(p+1)$ -asteinen ja ei-stokastinen $n \times (p+1)$ -matriisi.*

3.5 Logistinen regressio

Logistisen regression tapauksessa vastemuuttuja Y on kaksiluokkainen. Oletetaan tässä tutkielmassa, että sen mahdolliset arvot ovat vain nolla ja yksi - asema pitää hinnan samana tai muuttaa sitä. Oletetaan lisäksi, että

$$\begin{aligned} P(Y = 1 | \mathbf{X} = \mathbf{x}) &= \pi(\mathbf{x}) \text{ ja} \\ P(Y = 0 | \mathbf{X} = \mathbf{x}) &= 1 - \pi(\mathbf{x}). \end{aligned}$$

Tästä seuraa, että

$$\begin{aligned} E[Y|\mathbf{X} = \mathbf{x}] &= \pi(\mathbf{x}) \text{ ja} \\ \text{Var}[Y|\mathbf{X} = \mathbf{x}] &= \pi(\mathbf{x})(1 - \pi(\mathbf{x})). \end{aligned}$$

Koska todennäköisyydet ovat aina välillä $[0, 1]$, niin logistisella linkkifunktiolla kuvataan todennäköisyyksien joukko $(0, 1)$ reaaliakselille. Logistinen linkki voidaan kirjoittaa nyt muodossa

$$\text{logit}(\pi(\mathbf{x})) = \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \mathbf{x}^T \boldsymbol{\beta} = \eta.$$

Tästä puolestaan voidaan helposti johtaa selittävien muuttujien yhteys vasteeseen:

$$\frac{\pi(x_1, \dots, x_j + 1, \dots, x_p)}{1 - \pi(x_1, \dots, x_j + 1, \dots, x_p)} = e^{\beta_j} \frac{\pi(x_1, \dots, x_j, \dots, x_p)}{1 - \pi(x_1, \dots, x_j, \dots, x_p)}.$$

Tätä voidaan tulkita niin, että yhden yksikön kasvu selittävässä muuttujassa x_j on yhteydessä kerrannaisesti niin sanottuun vedonlyöntisuhteeseen $\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}$. Edeltä voidaan johtaa vielä varsinaiset todennäköisyydet:

$$\pi(\mathbf{x}) = \frac{e^{\mathbf{x}^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}^T \boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{x}^T \boldsymbol{\beta}}}.$$

Käydään seuraavaksi läpi logistisen regression SU-estimointiperiaate, koska se eroaa äskeisestä monen selittäjän lineaarisen regressiomallin estimoinnista. Kuten monen muuttujan lineaarisessa regressioanalyysissä, niin myös logistisessa regressiossa oletetaan vasteiden arvojen olevan ehdollisesti riippumattomia. Edellä todettiin, että $P(Y = 1|\mathbf{X} = \mathbf{x}) = \pi(\mathbf{x})$, jolloin siis $y_i \sim \text{Bernoulli}(\pi(\mathbf{x}_i^T))$ kaikilla $i = 1, \dots, n$. Täten uskottavuusfunktioksi saadaan

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i^T)^{y_i} [1 - \pi(\mathbf{x}_i^T)]^{1-y_i}.$$

Tämän logaritminuoto maksimointia varten puolestaan voidaan kirjoittaa muotoon

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \log(L(\boldsymbol{\beta})) = \sum_{i=1}^n \log[\pi(\mathbf{x}_i^T)^{y_i} (1 - \pi(\mathbf{x}_i^T))^{1-y_i}] \\ &= \sum_{i=1}^n [y_i \log(\pi(\mathbf{x}_i^T)) + (1 - y_i) \log(1 - \pi(\mathbf{x}_i^T))]. \end{aligned}$$

Uskottavuusfunktion suurin arvo löytyy nyt logaritmisesta uskottavuusfunktion osittaisderivaattojen nollakohdista, jotka ovat

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^n \left[y_i \frac{\partial}{\partial \beta_0} \log \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) + (1 - y_i) \frac{\partial}{\partial \beta_0} \log \left(1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) \right] \\
&= \sum_{i=1}^n \left[y_i \frac{e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} + (1 - y_i) \left(-1 + \frac{e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) \right] \\
&= \sum_{i=1}^n \left[y_i - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right] \text{ ja} \\
\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} &= \sum_{i=1}^n \left[y_i \frac{\partial}{\partial \beta_j} \log \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) + (1 - y_i) \frac{\partial}{\partial \beta_j} \log \left(1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) \right] \\
&= \sum_{i=1}^n \left[y_i x_{ij} \frac{e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} + (1 - y_i) x_{ij} \left(-1 + \frac{e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} \right) \right] \\
&= \sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{X}_i^{\mathbf{T}})],
\end{aligned}$$

missä $j = 1, \dots, p$.

Koska halutaan löytää, missä uskottavuusfunktio saa maksiminsa, niin asetetaan osittaisderivaatat nolliksi. Näin saadaan yhtälöt

$$\sum_{i=1}^n [y_i - \pi(\mathbf{X}_i^{\mathbf{T}})] = 0 \text{ ja } \sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{X}_i^{\mathbf{T}})] = 0.$$

Näistä yhtälöistä ei voida ratkaista kertoimia $\hat{\boldsymbol{\beta}}$ suljetussa muodossa. Tämän vuoksi estimointi suoritetaan loppuun numeerisella menetelmällä. Yksi tapa on esimerkiksi iteratiivinen painotettu pienimmän neliösumman menetelmä, joka kuitenkin sivutehtaan tässä tutkielmassa.

3.6 Yleistetty pienimmän neliösumman menetelmä

Pienimmän neliösumman menetelmässä mallin $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ jäännösten kovarianssimatriisiksi oletettiin $\sigma^2 \mathbf{I}$. Jos oletus jäännösten korreloimattomuudesta ei ole voimassa, pienimmän neliösumman menetelmän estimaattorit voivat olla tehottomia - varsinkin regressiokertoimien varianssi aliestimoituu. Tarkastellaan yleisempää tilannetta, jossa sallitaan jäännösten autokorrelaatio. Ilmiötä esiintyy erityisesti silloin, kun vastemuuttuja ja osa selittävistä muuttujista ovat aikasarjoja. Olkoon regressiomalli kuten edellä aliluvussa 3.1, mutta $\boldsymbol{\epsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{V})$ eli $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{V}$. Oletetaan nyt kuitenkin jäännösten homoskedastisuus ja oletetaan tunnetuksi positiivisesti definitti jäännösten kovarianssimatriisi

$$\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{V} = \sigma^2 \begin{pmatrix} 1 & \rho_{21} & \dots & \rho_{n1} \\ \rho_{21} & 1 & \dots & \rho_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}.$$

Koska \mathbf{V} oletettiin positiivisesti definiittiseksi matriisiksi, sille voidaan muodostaa spektraalihajotelma

$$\mathbf{V} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^T,$$

missä matriisin \mathbf{C} sarakkeet ovat matriisin \mathbf{V} karakteristiset vektorit ja matriisi $\boldsymbol{\Lambda}$ on diagonaalimatriisi, joka sisältää matriisin \mathbf{V} ominaisarvot. Määritellään matriisit $\boldsymbol{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ ja $\mathbf{T} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}$, jolloin $\mathbf{V} = \mathbf{T}\mathbf{T}^T$. Olkoon $\mathbf{P} = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}$, mistä seuraa että $\mathbf{V}^{-1} = \mathbf{P}\mathbf{P}^T$.

Kerrotaan yhtälö (3.2) puolittain vasemmalta matriisilla \mathbf{P} , jolloin saadaan yhtälö

$$\mathbf{P}\mathbf{Y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\epsilon}.$$

Merkitään tätä lyhyemmin

$$\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\epsilon}_*.$$

Nyt transformoitujen jäännösten kovarianssimatriisi on

$$\text{Cov}[\boldsymbol{\epsilon}_*] = E[\boldsymbol{\epsilon}_*\boldsymbol{\epsilon}_*^T] = \mathbf{P}\sigma^2\mathbf{V}\mathbf{P}^T = \sigma^2\mathbf{I}.$$

Tästä seuraa, että pienimmän neliösumman estimointimenetelmän oletus autokorrelaation puuttamisesta on voimassa ja menetelmää voidaan käyttää käyttäen transformoiduille muuttujille \mathbf{Y}_* ja \mathbf{X}_* .

Yleisetty pienimmän neliösumman estimaattori saadaan nyt muotoon

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{Y}_* = (\mathbf{X}^T\mathbf{P}^T\mathbf{P}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}^T\mathbf{P}\mathbf{Y} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}.$$

Estimaattorin kovarianssimatriisi lasketaan kuten pienimmän neliösumman menetelmässä, mutta nyt kuitenkin transformoiduilla muuttujilla \mathbf{Y}_* ja \mathbf{X}_* :

$$\text{Cov}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}_*^T\mathbf{X}_*)^{-1} = \sigma^2(\mathbf{X}^T\mathbf{V}\mathbf{X})^{-1}.$$

Edellä oletettiin jäännösten kovarianssimatriisi \mathbf{V} tunnetuksi. Todellisessa sovellustilanteessa tämä on kuitenkin harvinaista, koska matriisista \mathbf{V} on harvoin täsmällistä ennakkokäsitystä. Tarkastellaan seuraavaksi pääpiirteittäin iteratiivista Cochrane-Orcutt-estimointimenetelmää [4], jossa matriisi \mathbf{V} korvataan sen aineistosta lasketulla estimaatillaan. Aluksi todetaan, että jos matriisia \mathbf{V} ei ole rajoitettu mitenkään, estimoitavana on $n(n+1)/2$ autokorrelaatioparametria. Parametreja on enemmän kuin havaintojen määrä n ja mahdotonta estimoida aineiston avulla. Tämän takia kovarianssirakenteesta on tehtävä ennalta rajoituksia. Tähän on vaihtoehtoisia tapoja.

Voidaan esimerkiksi asettaa autokorrelaatio nolaksi jostain viiveestä q ($q < n(n+1)/2$) lähtien. Olkoon tuntemattomat autokorrelaatiot vektorin $\boldsymbol{\theta}$ alkioita, jolloin $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$. Siirrytään nyt käyttämään kovarianssimatriisia $\widehat{\mathbf{V}} = \mathbf{V}(\widehat{\boldsymbol{\theta}})$ oikean tuntemattoman populaation korrelaatiomatriisin \mathbf{V} sijasta.

Olkoon malli ja sen jäännösten autokorrelaatorakenne seuraava:

$$\begin{aligned} (1) \quad y_i &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, \\ (2) \quad \epsilon_i &= \rho_1 \epsilon_{i-1} + \dots + \rho_q \epsilon_{i-q} + u_t, \quad u_t \sim N(0, \sigma_u^2). \end{aligned}$$

Iteratiivinen Cochrane–Orcutt-estimointialgoritmi vektorin $\boldsymbol{\beta}$ estimointiin on pääpiirteittäin seuraava:

1. Estimoidaan malli (1) ensin pienimmän neliösumman menetelmällä.
2. Estimoidaan otosautokorrelaatiokertoimet $\widehat{\rho}_1, \dots, \widehat{\rho}_q$ mallista (1).
3. Sovelletaan pienimmän neliösumman menetelmää malliin:

$$\begin{aligned} \left(y_i - \sum_{l=1}^q \widehat{\rho}_l y_{i-l} \right) &= \beta_0 \left(1 - \sum_{l=1}^q \widehat{\rho}_l \right) + \beta_1 \left(x_{1i} - \sum_{l=1}^q \widehat{\rho}_l x_{1(i-l)} \right) + \\ &\dots + \beta_p \left(x_{pi} - \sum_{l=1}^q \widehat{\rho}_l x_{p(i-l)} \right) + \left(u_i - \sum_{l=1}^q \epsilon_{i-l} \right). \end{aligned}$$

Korvataan malli (1) tällä mallilla. Toistetaan vuorotellen vaiheita 2. ja 3., kunnes saavutetaan riittävä konvergenssi.

Tässä tutkielmassa ei käsitellä tätä estimointialgoritmia tämän yksityiskohtaisemmalta tasolla.

4 Tilastollinen mallintaminen ja tulokset

Tässä luvussa esitellään lyhyesti empiirisiä menetelmiä Johdanto-kappaleen tutkimuskysymysten 1-3 (sivulla 3) vastaamiseen. Analyysit perustuvat avoimiin datoihin (kappale 2). Polttoainehinnoissa on puuttuvia havaintoja, mutta tulosten voidaan ajatella kuvaavan tilannetta helposti saatavilla olevan tiedon perusteella. On siis käytetty näkymää, johon kenellä tahansa on pääsy Internetin kautta. Puuttuvan datan ja kuluttajien lisäämien hintojen vuoksi datan avulla tehdyt johtopäätökset todellisista hintailmiöistä voivat jossain määrin olla kyseenalaisia, mutta oikeastikaan kuluttajilla ei ole parempaa kvantitatiivista näkymää tästä ilmiöstä. Lähemmäs totuutta veisi

myös tiedot asemien ja jalostamoiden varastotasoista, katteista ja volyyymeista. Lisäksi huoltoasemien luokittelu eri tyyppeihin kuten suurin matkailuasemiin ja pieniin kylmäasemiin voisi tuoda lisäarvoa. Nyt käytetty aineisto on data-analyysin nykyhengen mukaista, ainakin lähellä niin sanottua 'Big dataa'. Aineistoa kertyy mittariken-tiltä suuret määrät tietokantoihin muiden nähtäväksi, ja tätä dataa voidaan edelleen rikastaa ja analysoida. Pro gradun yksi kulmakivistä onkin tutkia, miten avointa ja massiivista dataa voidaan hyödyntää. Tilastollinen mallintaminen on tehty SAS[®]-ohjelmiston versiolla 9.3.

4.1 Tutkimuskysymys 1: Miten ennustaa tietyn aseman seuraavan päivän hintaa?

Aineiston ja ennakkotietojen mukaan polttoaineasemat pitävät monesti usean päivän ajan samaa hintaa eri polttoainelaaduille. Aineiston mukaan esimerkiksi polttoaine-laadun 95E hintaa pidettiin keskimäärin samana noin viisi päivää. Siksi mallinnuskin tehdään kaksivaiheisesti: aluksi selvitetään, milloin ja mistä syystä hintaa lähdetään muuttamaan ja sen jälkeen mallinnetaan uusi hinnoittelu. Mallinnuksella selvitetään kvantitatiivisesti, miten aseman oma aikaisempi hinnoittelu ja lähialueen hinnat vaikuttavat uuteen hintaan. Jos mallintaminen tehtäisiin yhdellä kerralla lineaarisilla malleilla, niin tuloksena olisi pieni hintamuutos melkein jokaiselle päivälle. Tämä ei vastaa tositilannetta. Polttoaineiden hinnat on jälkimmäisessä mallinnusvaiheessa laatukohtaisesti keskitetty polttoaine.net:n keskihintojen ympärille. Tällöin vaste-muuttujana on poikkeama päiväkeskiarvosta, jolloin muun muassa globaalit hintaan vaikuttavat tekijät saadaan kontrolloitua ja päästään tutkimaan paikallista vähittäis-kaupan hintakilpailua. Kuluttajat voivat ajatella hintojen olevan luonnostaan tietyllä tasolla, ja heitä kiinnostaa miten alueen hinnat muuttuvat toistensa suhteen lyhyellä aikavälillä. Lähdeaineistosivustolla <http://www.polttoaine.net> raportoidaan myös Suomen keskihinnat.

4.1.1 Hinnanmuuttamispäätöksen mallinnus

Mallinnus

Tarkastellaan aluksi, mitkä tekijät selittävät aseman päätöstä lähteä muuttamaan hintaa. Ennakkokäsitysten ja saatavilla olevan aineiston perusteella aseman oma hintahistoria ainakin on potentiaalinen tekijä: jos hintaa on juuri muutettu, sitä ei välttämättä vähään aikaan haluta muuttaa uudelleen. Toisaalta taas aseman on sopeutettava hintansa vastaamaan markkina-alueensa hintatasoa: lähimmän kilpailijan eli lähimmän aseman hintamuutoksia on syytä seurata, koska monesti autoilijat näkevät aseman läheisyydessä muidenkin asemien hintoja. Lisäksi markkina-alueen hintata-son muutoksilla voidaan olettaa olevan myös vaikutusta aseman hinnanmuutospäätökseen. Tätä kuvataan kymmenen kilometrin säteellä olevalla halvimmalla hinnalla. Oletetaan vielä, että hinnanmuutoksiin voi vaikuttaa viikonpäivä: esimerkiksi viikon-loppuisin mahdollisesti asetetaan systemaattisesti eritasoinen hinta kuin arkipäiville.

Rakennetaan aseman muutospäätökselle polttoainekohtaisesti logistiset regressiomallit, joissa hinnanmuutospäätöstä selitetään aseman samojen hintojen kestolla, aseman edellisen päivän hinnan poikkeamasta lähimmän aseman uusimmasta hinnasta ja poikkeamasta 10 kilometrin säteen halvimmasta hinnasta sekä viikonpäivällä. Ajatellaan hintapoikkeamat kilpailijoista vielä itseisarvona, mikä tarkoittaa, että ei oteta kantaa siihen onko asemalla kalliimpaa tai halvempaa kuin lähiympäristössä: hintamuutospäätös on symmetristä oli aseman hintataso kummassa päässä tahansa. Kun valitaan logistiseen regressioon logit-linkkifunktio, niin malli voidaan kirjoittaa muotoon:

(4.1)

$$\log \frac{P(M_{ijt} = 1)}{1 - P(M_{ijt} = 1)} = \beta_{M,i} + \beta_{\text{sama},i} x_{\text{sama},ijt} + \beta_{\text{Mlähin},i} |x_{\text{lähin},ijt} - y_{ij(t-1)}| \\ + \beta_{\text{Mmin},i} |x_{\text{min},ijt} - y_{ij(t-1)}| + \beta_{\text{ti},i} p_{\text{ti},t} + \dots + \beta_{\text{su},i} p_{\text{su},t} + \epsilon_{ijt},$$

missä

i	viittaa polttoainelaatuihin ($i = 95\text{E}, 98\text{E}, \text{DI}$);
j	viittaa huoltoasemaan ($j = \text{AsemaID1}, \dots, \text{AsemaID723}$);
t	viittaa päivään (1.1.2010, ..., 2.10.2012);
M_{ijt}	on indikaattorimuuttuja hintamuutokselle: $M_{ijt} = 1$, jos $y_{ijt} = y_{ij(t-1)}$
y_{ijt}	on polttoainelaadun i (€/l) myyntihinta asemalla j päivänä t ; muuten $M_{ijt} = 0$;
$x_{\text{sama},ijt}$	kuvaa, kuinka monta päivää polttoainelaadun i aseman j hinta on ollut sama päivänä t ;
$x_{\text{lähin},ijt}$	on polttoainelaadun i hinta aseman j lähimmällä asemalla päivänä t ;
$x_{\text{min},ijt}$	on polttoainelaadun i minimihinta 10 km:n säteellä asemasta j päivänä t ;
p	viittaa viikonpäivään: $p_{\text{ti},t} = 1$, jos viikonpäivä päivänä t on tiistai, $p_{\text{ti},t} = 0$ muuten; loput viikonpäivävaihtoehdot vastaavasti;
ϵ_{ijt}	on virhetermi ja oletetaan, että $\epsilon_{ijt} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

Estimoidaan mallin (4.1) parametrit $\beta_{M,i}$, $\beta_{\text{sama},i}$, $\beta_{\text{Mlähin},i}$, $\beta_{\text{Mmin},i}$, $\beta_{\text{ti},i}$, ..., $\beta_{\text{su},i}$ SAS-ohjelmiston LOGISTIC-proseduurilla. Tällöin siis estimoidaan kolme mallia eli jokainen polttoainelaatu saa oman mallinsa. Kun näin tehtiin, niin alueen minimihintojen $\beta_{\text{Mmin},i}$ kerrointen merkiksi tuli negatiivista, mikä on oletusten vastaista. Tällöin kilpailijan suurempi poikkeama kilpailijoista olisi yhteydessä suurempaan todennäköisyyteen pitää hintaa muuttumattomana. Oletuksena oli, että suuri ero kurottaisiin kiinni uudella hinnalla. Syy tähän negatiiviseen etumerkkiin on varmaankin lähimmän hinnan ja kymmenen kilometrin säteen minimihinnan välinen positiivinen korrelaatio ($r^2 = 0.5$). Poistetaan kertoimet $\beta_{\text{Mmin},i}$ jokaisen polttoainelaadun mallista, jotta lähimmän aseman hinnan kerroin estimoituisi tarkemmin. Malli voidaan näin ollen kirjoittaa muotoon

(4.2)

$$\log \frac{P(M_{ijt} = 1)}{1 - P(M_{ijt} = 1)} = \beta_{M,i} + \beta_{sama,i} x_{sama,ijt} + \beta_{Mlähin,i} |x_{lähin,ijt} - y_{ij(t-1)}| \\ + \beta_{ti,i} p_{ti,t} + \dots + \beta_{su,i} p_{su,t} + \epsilon_{ijt}.$$

Estimointitulokset malleille (4.2) on esitetty polttoainelaadittain taulukoissa 2, 3 ja 4. Viikonpäivissä maantantai on referenssitasona. Muutos on indikaattorimuuttuja, jonka perustaso on 'hintaa ei muuteta'.

Taulukko 2: Mallin (4.1) estimoinnin tulokset polttoainelaadulle 95E.

Parametri	Estimaatti	Keskivirhe	<i>p</i> -arvo
$\beta_{M,B95}$	-1.96	0.02	< 0.001
$\beta_{sama,B95}$	-0.02	0.02	< 0.001
$\beta_{Mlähin,B95}$	4.47	0.33	< 0.001
$\beta_{ti,B95}$	0.57	0.03	< 0.001
$\beta_{ke,B95}$	0.33	0.03	< 0.001
$\beta_{to,B95}$	0.24	0.03	< 0.001
$\beta_{pe,B95}$	0.29	0.03	< 0.001
$\beta_{la,B95}$	0.08	0.03	< 0.01
$\beta_{su,B95}$	-0.43	0.03	< 0.001

Taulukko 3: Mallin (4.1) estimoinnin tulokset polttoainelaadulle 98E.

Parametri	Estimaatti	Keskivirhe	<i>p</i> -arvo
$\beta_{M,B95}$	-1.83	0.02	< 0.001
$\beta_{sama,B95}$	-0.02	0.02	< 0.001
$\beta_{Mlähin,B95}$	1.84	0.19	< 0.001
$\beta_{ti,B95}$	0.55	0.03	< 0.001
$\beta_{ke,B95}$	0.35	0.03	< 0.001
$\beta_{to,B95}$	0.28	0.03	< 0.001
$\beta_{pe,B95}$	0.31	0.03	< 0.001
$\beta_{la,B95}$	0.13	0.03	< 0.001
$\beta_{su,B95}$	-0.39	0.03	< 0.001

Tulosten tulkinta

Malliyhtälöstä (4.2) ja sen estimointituloksista eli Taulukosta 2, 3 ja 4 voidaan johdattaa sovelluttuja tuloksia tutkimuskysymysten vastaamiseen. Aloitetaan tulkinnat ve-

Taulukko 4: Mallin (4.1) estimoinnin tulokset polttoainelaadulle diesel.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{M,B95}$	-1.93	0.024	< 0.001
$\beta_{sama,B95}$	-0.02	0.002	< 0.001
$\beta_{Mlähin,B95}$	4.62	0.25	< 0.001
$\beta_{ti,B95}$	0.57	0.03	< 0.001
$\beta_{ke,B95}$	0.37	0.03	< 0.001
$\beta_{to,B95}$	0.29	0.03	< 0.001
$\beta_{pe,B95}$	0.32	0.03	< 0.001
$\beta_{la,B95}$	0.07	0.03	0.03
$\beta_{su,B95}$	-0.35	0.03	< 0.001

donlyöntisuhteista eli muuttamisen todennäköisyyden ja samana pitämisen todennäköisyyden suhteesta. Yhden yksikön muutos selittävässä muuttujassa aiheuttaa vedonlyöntisuhteeseen kerrannaisen vaikutuksen siten, että suhde muuttuu kyseisen β -kertoimen eksponentin verran (e^{β_j}). Nyt $\hat{\beta}_{Mlähin,B95} = 4.47$, joten jos kyseisen aseman lähin naapuriasema muuttaa hintaansa yhdellä sentillä suuntaan tai toiseen, niin 'riski' tehdä uudelleenhinnoittelu kasvaa $e^{\frac{4.47}{100}} = 1.046$ -kertaiseksi. Toisin sanoen jokainen sentti, mitä naapuriaseman 95E:n hinta lähtee eroamaan kyseisen aseman hinnasta on yhteydessä noin viiden prosentin riskiin asettaa uusi hinta bensiinille 95E. Luottamusvälin leveys tälle on [4.2, 5.0] prosenttiyksikköä. Kesto, jonka nykyinen hinta on ollut voimassa, ei ole suuremmin yhteydessä hintamuutospäätökseen: yksi lisäpäivä on yhteydessä 1.5 prosentin laskuun vedonlyöntisuhteessa. Viikonpäivien indikaattorimuuttujien kertoimet ovat myös tilastollisesti merkittäviä 95 %:n luottamustasolla. Tätä edesauttaa paljon suuri otoskoko: $n = 230000$. Aineistoa on hieman kutistettu poistamalla äärimmäisiä havaintoja - lähinnä kyseenalaisen pitkiä, jopa kahden kuukauden muuttumattomia hintoja. Poistolla ei ole kuitenkaan käytännön vaikutusta isoon kuvaan. Tarkemmat tulokset tästä sivuutetaan. Kaikkien muiden viikonpäivien kertoimet sunnuntaita lukuunottamatta ovat merkittävästi positiivisia. Tämä tarkoittaa, että sunnuntaisin ja maanantaisin hinnan muuttaminen on mallin mukaan epätodennäköisintä. Esimerkiksi jos muut mallin tekijät pidetään muuttumattomina, niin maanantaista tiistaihin siirryttäessä vedonlyöntisuhde $e^{0.57} = 1.77$ -kertaistuu. Luottamusvälin leveys tälle on 0.10 yksikköä suuntaansa. Polttoainelaatujen välillä kertoimet ovat kutakuinkin samanlaisia. Eniten esiin nousee kuitenkin bensiinilaadun 98E kerroin $\hat{\beta}_{Mlähin,B95} = 1.84$. Tämä on alle puolta pienempi kuin muilla laaduilla. Tämä viittaa B98:n hinnan olevan stabiilimpi kuin B95:n ja dieselin lähimmän aseman hinnoittelulle. Tarkempaa syytä on vaikea todentaa. Vastaava ristitulo suhde kuin edellä 95E:lle (1.05) on 98E:lle nyt 1.02.

Mallin antamat todennäköisyydet hintamuutospäätökselle on myös helppo johtaa yhtälöstä (4.1) tässä logistisen mallin tapauksessa. Jos esimerkiksi asema on pitänyt maanantaista asti kaksi päivää jo samaa hintaa 95E:lle ja jos lähimmän aseman hinta laskee 0.05 €/l, niin mallin mukaan hintaa tullaan keskiviikkona muuttamaan toden-

näköisyydellä

$$\frac{e^{-1.96-0.02*2+4.47*0.05+0.33*1}}{1 + e^{-1.96-0.02*2+4.47*0.05+0.33*1}} \approx 18.6\%.$$

Luottamusväli tälle on [18.5,18.7] %. Jos taas edellä mainittu viiden sentin litrahintamuutos markkinaalueella jätettäisiin tekemättä, niin mallin mukaan todennäköisyys hintamuutokselle tipahtaa 15.4 prosenttiin. Saadut todennäköisyydet ovat ehkä aika pienehköjä, mikä voi viitata jonkin oleellisen selittäjän puutteesta mallissa. Tällaisia voivat olla esimerkiksi aseman, tukun ja jalostajan varastotasot tai kateprosentin pakottaminen tietylle tasolle. Näihin tässä työssä ei ole mahdollista päästä sen tarkemmin käsiksi. Toisaalta ilmiö voi johtua myös sattumasta tai epäsystemaattisesta toiminnasta. Aineistoa tutkimalla tiettyä kaavaa ei suoraan nopeasti löydy eli hintoja muutetaan 'miten milloinkin'-periaatteella huolimatta kilpailijoista tai aseman omasta hintahistoriasta. Yksi jatkotarkastelumahdollisuus olisi haastatella alan työntekijöitä, ymmärtää polttoainemarkkinoiden liiketoimintaa paremmin ja kerätä enemmän dataa asian tutkimiseksi tarkemmin.

4.1.2 Uuden hinnan mallinnus

Tarkastellaan seuraavaksi tilannetta, jossa asema päättää muuttaa polttoainelaadun hintaa. Kiinnostavaa on tietää, paljonko uusi hinta on. Lisäksi tämä vastaa tutkimuskysymyksiin hintakilpailun osalta. Toisin sanoen miten hinnoittelu on kytköksissä markkina-alueen hintatasoon. Ajatellaan malliin polttoainekohtaisesti potentiaalisiksi selittäviksi muuttujiksi tietyn aseman ja tietyn päivän hinnalle aseman edellistä hintaa, alueen minimihintaa ja lähimmän aseman hintaa. Aseman edellisessä hinnassa voidaan olettaa olevan paras tieto seuraavan hinnan tasosta. Monesti halvat asemat pysyvät halpoina ja kalliit kalliina. Kilpailijoiden muutoksilla taas voidaan ajatella olevan vaikutus muutoksen suuntaan. Keskistetään hinnat vielä Suomen keskihintojen ympärille. Nämä keskihinnat kuluttujankin on helppo saada selville esimerkiksi lähdesivustolta [11]. Tällöin hinnat ovat poikkeamia keskihinnoista. Tämä on mallintamiselle suotuista ajatellen kilpailutekijöiden yhtettä hinnoitteluun: aikasarjadimensio vähenee ainakin osittain ja regressiomallilla päästään tehokkaammin käsiksi lyhyen aikavälin alueelliseen hinnoitteluun markkina-alueen sisällä. Polttoainneiden hintojen yleistason voidaan ajatella muodostuvan luonnostaan, tässä kohtaa ei puututa siihen. Tästä seuraa toki oletus, että kilpailuun vastaaminen on samantyyppistä yli ajan, mikä ei kuulosta intuitiivisesti mahdottomalta. Aineisto on asema-, päivä- ja polttoainelaatutasoista. Rakennetaan polttoainelaaduttain uusille hinnoille y_{ijt} regressiomallit:

(4.3)

$$(y_{ijt} - \bar{y}_{it}) = \beta_{0,i} + \beta_{AR1,i} (y_{ij(t-1)} - \bar{y}_{i(t-1)}) + \beta_{\min,i} (x_{\min,it} - \bar{y}_{it}) + \beta_{\text{lähin},i} (x_{\text{lähin},it} - \bar{y}_{it}) + \epsilon_{ijt},$$

missä

i	viittaa polttoainelaatuihin ($i = 95E, 98E, DI$);
j	viittaa huoltoasemaan ($i = 1, \dots, 723$);
t	viittaa päivään (1.1.2010, ..., 2.10.2012);
y_{ijt}	on polttoainelaadun i (€/l) myyntihinta asemalla j päivänä t ;
\bar{y}_{it}	on polttoainelaadun i Suomen keskihinta päivänä t [11];
$x_{AR1,ijt}$	on polttoainelaadun i myyntihinta asemalla j päivänä ($t-1$);
$x_{\min,ijt}$	on polttoainelaadun i minimihinta 10 km:n säteellä asemasta j päivänä t ;
$x_{\text{lähin},ijt}$	on polttoainelaadun i hinta aseman j lähimmällä asemalla päivänä t ;
ϵ_{ijt}	on virhetermi ja oletetaan, että $\epsilon_{ijt} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

Estimoidaan SAS-ohjelmiston REG-proseduurilla mallin (4.3) parametrit eli β -kertoimet $\beta_{0,i}$, $\beta_{AR1,i}$, $\beta_{\min,i}$, $\beta_{\text{lähin},i}$ ja σ^2 . Jätetään aineistoon polttoainelaatukohtaisesti vain kaikki ne havainnot, joissa asema on muuttanut hintaansa edellisestä päivästä. Tämä on siis eri asia kuin muutos poikkeamasta Suomen keskihinnasta. Bensiinilaadun 95E tapauksessa havaintomäärä on 52225, 98E-laadulle 41597 ja dieselin tapauksessa havainnot on 43843. Kaiken kolmen mallin eli jokaisen polttoainelaadun kohdalla estimointituloksista käy ilmi, että $\beta_{\min,i}$ ei ole missään tilastollisesti merkitsevä. Kerroin jää olemattoman pieneksi, joten se poistetaan malleista. Karsitut mallit saavat nyt muodon

$$(4.4) \quad (y_{ijt} - \bar{y}_{it}) = \beta_{0,i} + \beta_{AR1,i} (y_{ij(t-1)} - \bar{y}_{i(t-1)}) + \beta_{\text{lähin},i} (x_{\text{lähin},it} - \bar{y}_{it}) + \epsilon_{ijt}.$$

Mallin (4.4) estimointitulokset on esitetty polttoainelaaduittain taulukoissa 5, 6 ja 7. Kuvissa 2 – 6 on esitetty mallin (4.4) diagnostiikkakuvioita. Kuvassa 2 on polttoainelaaduittain sovitteiden ja jäännösten hajontakuvio, Kuvassa 3 standardoitujen jäännösten histogrammi, Kuvassa 4 selittävän muuttujan $(y_{ij(t-1)} - \bar{y}_{i(t-1)})$ ja jäännösten hajontakuvio, Kuvassa 5 selittävän muuttujan $(x_{\text{lähin},it} - \bar{y}_{it})$ ja jäännösten hajontakuvio ja Kuvassa 6 on viiden satunnaisesti valitun aseman jäännösten aikasarja autokorrelaation kuvaamiseksi. Kuten kuvista näkyy, sovitteita vastaavat jäännökset ovat jakautuneet satunnaisesti eri sovitteiden arvoilla ja kuviossa ei esiinny liikaa systematiikkaa. Histogrammitkin ovat likimain normaalijakaumaa noudattavia. Selit-

Taulukko 5: Mallin (4.4) estimoinnin tulokset polttoainelaadulle 95E.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,B95}$	0.00	< 0.001	< 0.001
$\beta_{AR1,B95}$	0.66	0.004	< 0.001
$\beta_{\text{lähin},B95}$	0.13	0.004	< 0.001

Selitysaste 56 %.

Taulukko 6: Mallin (4.4) estimoinnin tulokset polttoainelaadulle 98E.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,B98}$	0.00	< 0.001	0.02
$\beta_{AR1,B98}$	0.78	0.004	< 0.001
$\beta_{\text{lähin},B98}$	0.06	0.004	< 0.001

Selitysaste 67 %.

Taulukko 7: Mallin (4.4) estimoinnin tulokset polttoainelaadulle diesel.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,DI}$	0.00	< 0.001	0.32
$\beta_{AR1,DI}$	0.64	0.005	< 0.001
$\beta_{\text{lähin},DI}$	0.15	0.005	< 0.001

Selitysaste 55 %.

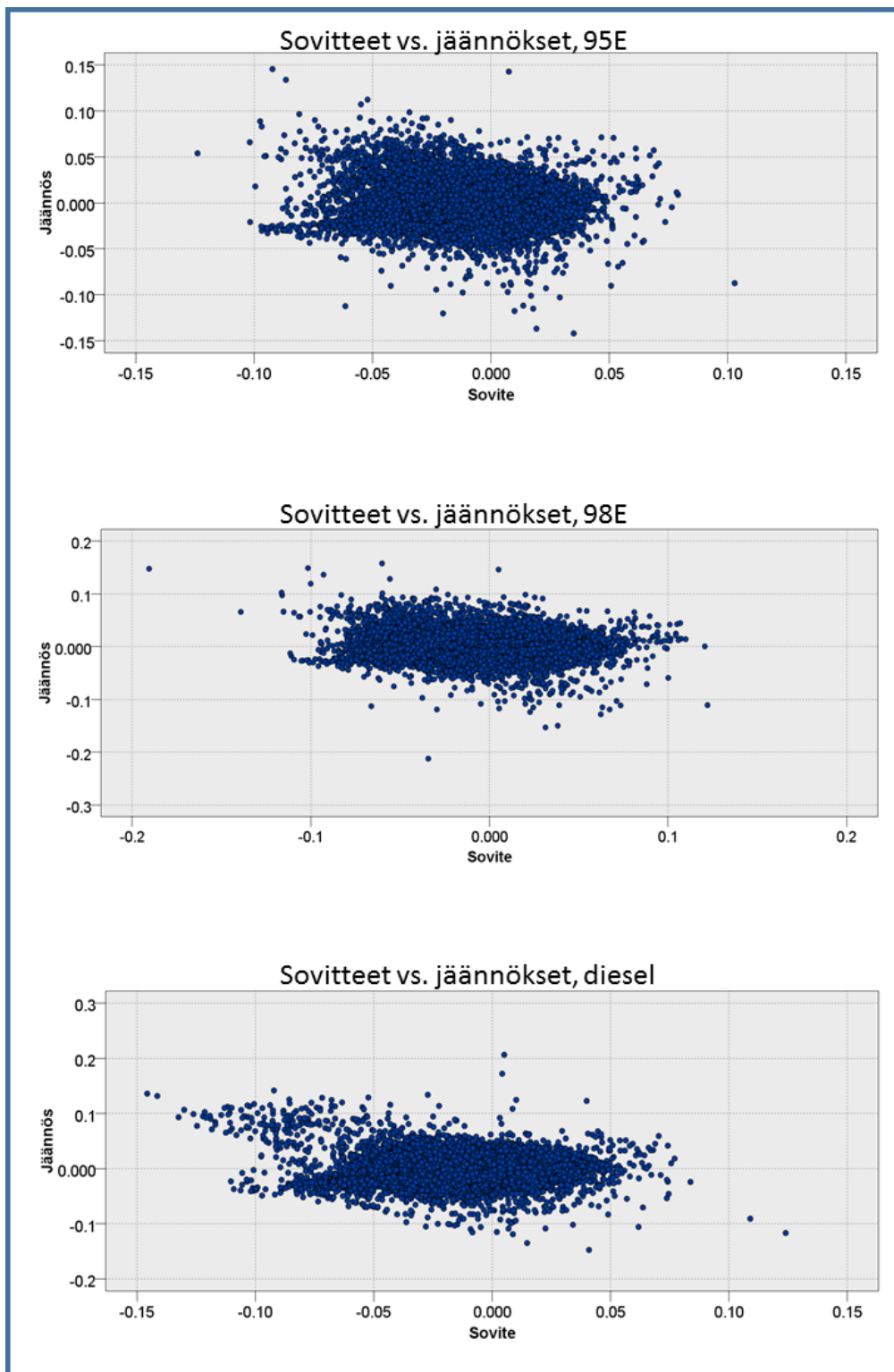
tävät muuttujatkaan eivät näyttäisi korreloivan jäännösten kanssa. Myöskään suurta silmään pistävää autokorrelaatiota ei ole nähtävissä. Voidaan siis todeta, että lineaarinen malli sopii mallintamaan tilannetta hyvin. Selitysasteetkin 55 % – 67 % ovat ennustusmielessä tyydyttäviä.

Tulosten tulkinta

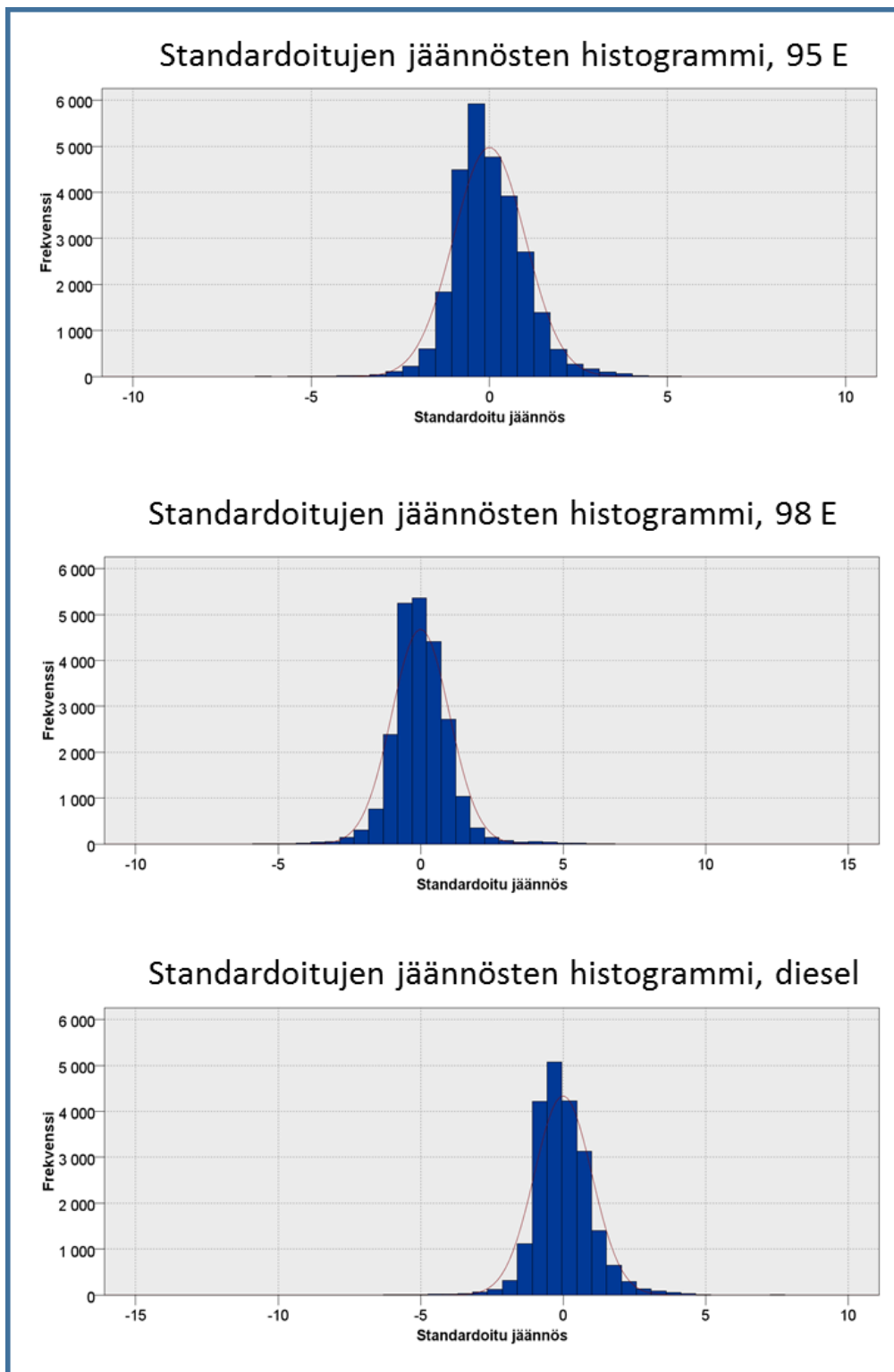
Tutkitaan seuraavaksi toisen vaiheen mallinnustuloksia. Nyt siis tarkastellaan tilannetta, että asemalla on päätetty asettaa uusi hinta. Katsotaan, miten hinta muodostuu mallien kautta. Malliyhtälöstä (4.4) ja Taulukoista 5 seuraa, että polttoainelaadulle 95E ennustettu uusi hinta asemalle j päivänä t on

$$y_{B95,jt} = 0.00 + 0.66 (y_{ij(t-1)} - \bar{y}_{i(t-1)}) + 0.13 (x_{\text{lähin},it} - \bar{y}_{it}) + \bar{y}_{it}.$$

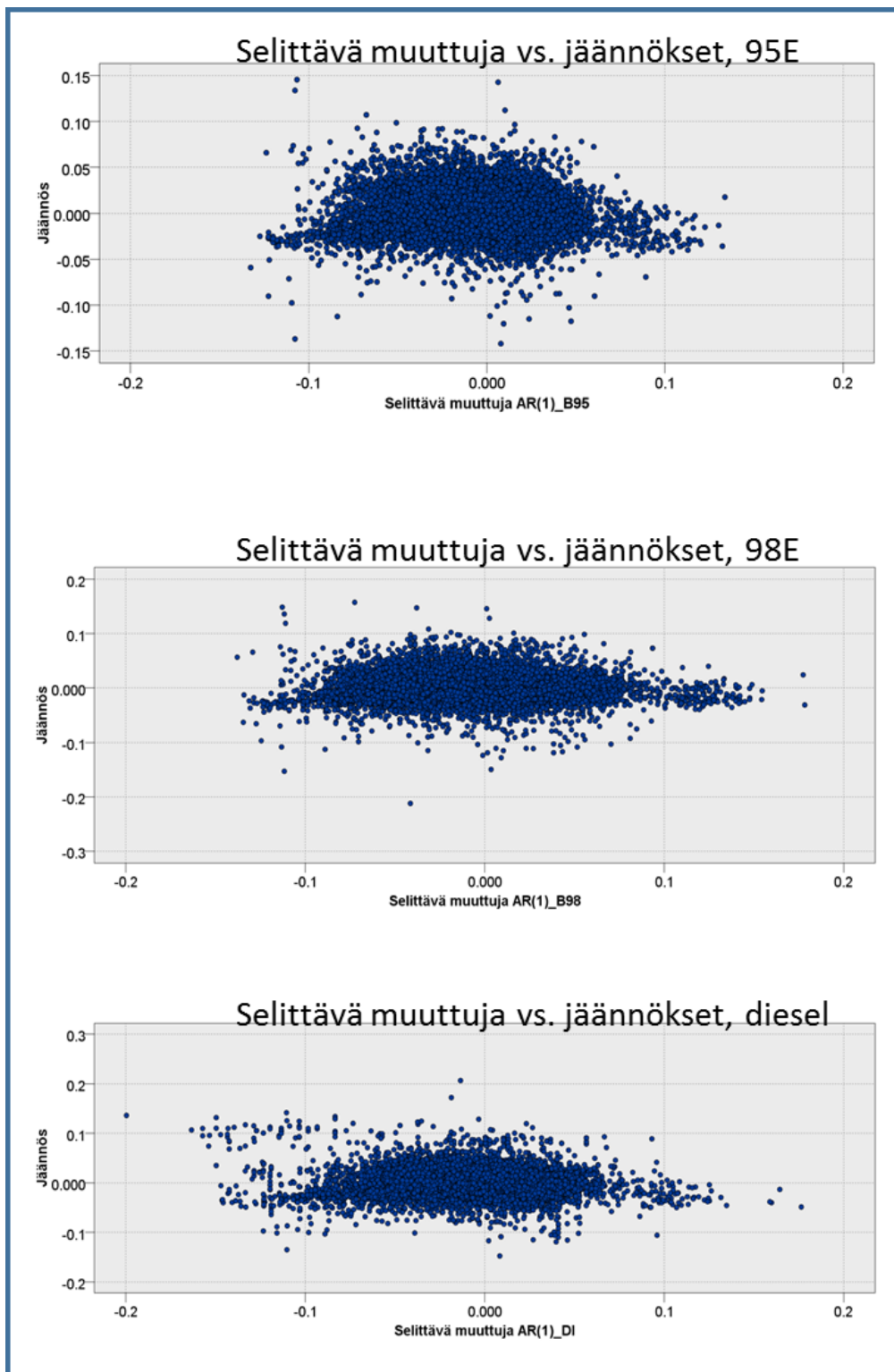
Tämä tarkoittaa polttoainelaadun 95E kohdalla, että kun muut mallin tekijät pysyvät ennallaan, niin asemalla hintaa muutetaan keskimäärin kertoimella 0.66 siihen suuntaan kuin sen edellinen hinta poikkesi Suomen keskihinnasta. Edelleen asema muuttaa hintaansa keskimäärin kertoimella 0.13 samaan suuntaan kuin kilpailija poikkeaa Suomen keskihinnasta. Paras ja tärkein tieto aseman uudelle hinnalle on täten sen omassa hintahistoriassa. Tämä tarkoittaa sitä, että asemalla on keskimäärin melko vakaa ja tietty oma hintatasonsa yleiseen hintatasoon suhteutettuna. Tällöin asema asettaa uuden hinnan lähinnä sen perusteella, miten sillä on ollut tapana hinnoitella. Monilla huoltoasemilla hinta tulee ketjulta annettuna, ja tässä saattaa vaikuttaa paljon paikallisesta kilpailusta riippumaton ketjutason tahtotila: millä hinnalla pystytään ja on järkevää myydä. Ympäröivät hinnat eli paikallinen kilpailu vaikuttaa kuitenkin kohdallisesti. Asemakohtaista vaihtelua toki esiintyy: osalla asemista hinnat joustavat



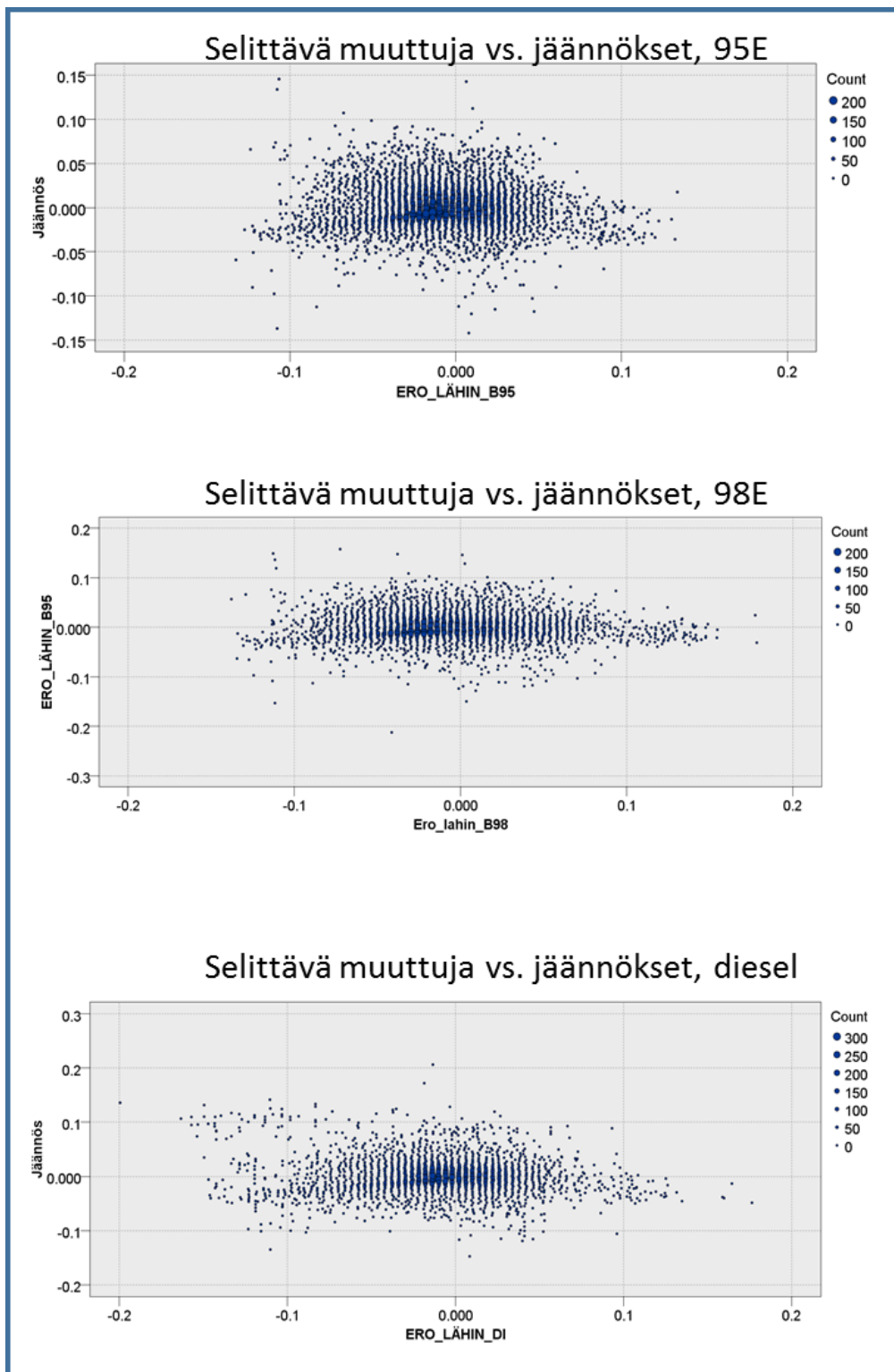
Kuva 2: Mallin (4.4) sovitteiden ja jäännösten hajontakuviot polttoainelaaduittain: 95E (ylh.), 98E (kesk.) ja diesel (alh.).



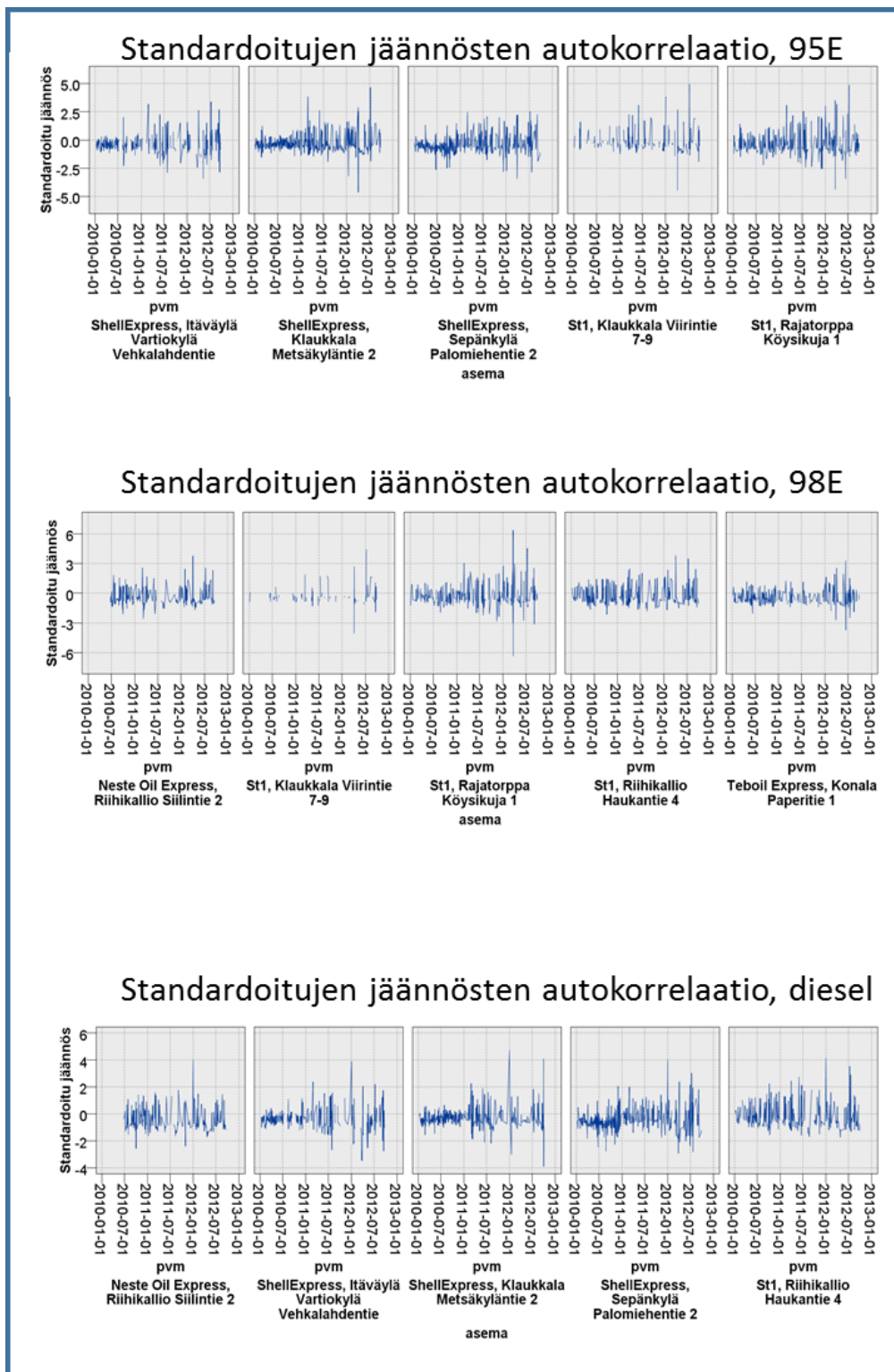
Kuva 3: Mallin (4.4) standardoitujen jäännösten histogrammi polttoainelaaduittain: 95E (ylh.), 98E (kesk.) ja diesel (alh.).



Kuva 4: Mallin (4.4) selittävän muuttujan ($y_{ij(t-1)} - \bar{y}_{i(t-1)}$) ja jäännösten hajontakuviot polttoainelaaduittain: 95E (ylh.), 98E (kesk.) ja diesel (alh.).



Kuva 5: Mallin (4.4) selittävän muuttujan ($x_{\text{lähin},it} - \bar{y}_{it}$) ja jäännösten hajontakuvio polttoainelaaduittain: 95E (ylh.), 98E (kesk.) ja diesel (alh.).



Kuva 6: Mallin (4.4) jäännösten aikasarja viideltä valitulta huoltoasemalta: 95E (ylh.), 98E (kesk.) ja diesel (alh.).

enemmän kilpailutilanteen mukaan, kun taas joillakin asemilla hinnat ovat jäykkiä markkina-alueen hintamuutoksille. Mallin selitysaste on 55 prosenttia. Tätä voidaan pitää kohtalaisen hyvänä ennustumielessä. Parametriestimaattien keskivirheiden ollessa hyvin pieniä kertoimia voidaan pitää käytännössä melko tarkkoina. Tämä selittyy mallin hyvällä osuvuudella, mutta vielä suurempana syynä tässä on suuri aineisto. Suuri aineisto mahdollistaa tarkemmat arviot kilpailutekijöiden vaikutukselle. Vakiotermin arvoksi estimointi antaa 95E:n tapauksessa 0.00. Tässä vaikuttaa mallin rakennuksessa valittu termien skaalaus. Koska nyt vakion arvo on nolla kaikissa polttoainelaaduissa, niin mallin voidaan ajatella muodostuvan kahdesta komponentista: aseman oma hinnoittelutyö sekä lähimmän aseman hintaan reagointi.

Myös dieselin kohdalla kertoimet ovat samansuuntaisia kuin bensiinilaadun 95E kohdalla, erot kertoimissa ovat parin sadasosan suuruisia. Bensiinilaadun 98E kohdalla puolestaan aseman oma hinnoittelu nousee vielä enemmän esiin (kerroin 0.78). Lähimmän hinnan kerroin on taas pienempi (0.06). Tämä viittaa aseman omapäisempään hinnoitteluun, ehkä 98E:n kohdalla hintasotaan ei ole tarvetta tai halua ryhtyä. Tämän laadun osuus asemien myynnistä lienee pienintä.

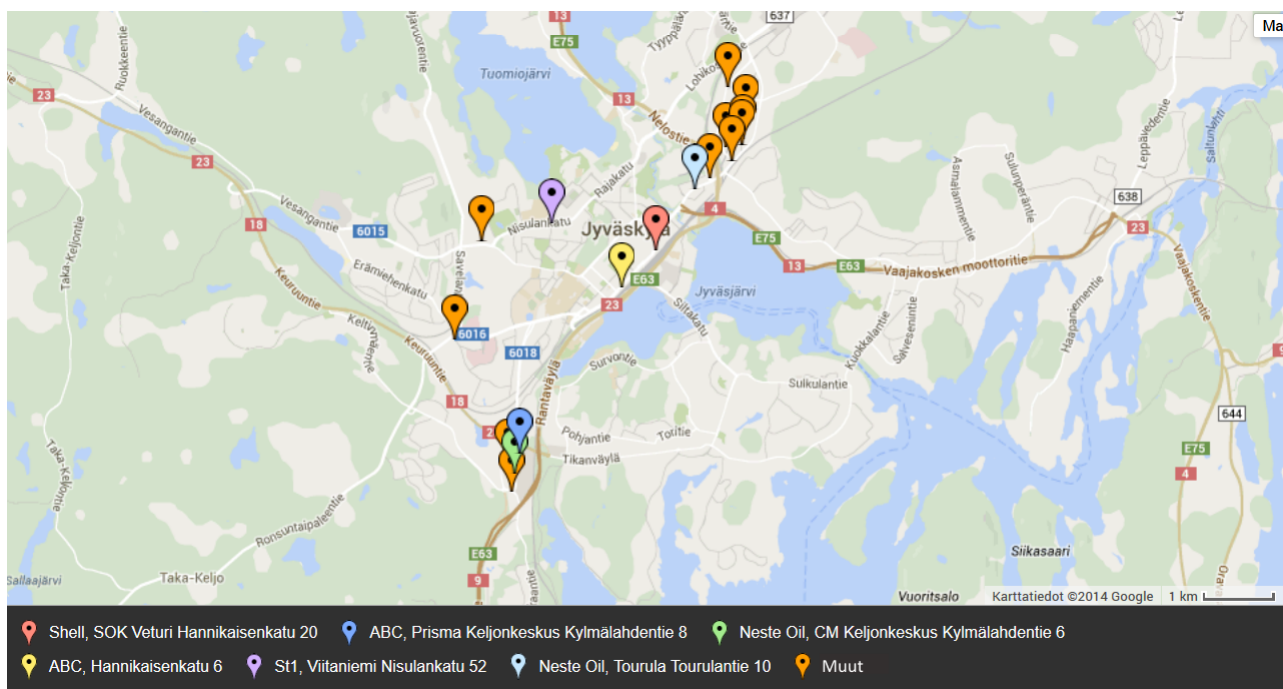
Tutkitaan seuraavaksi, miten koko mallia käytetään ennustumielessä. Apuna konkreettina antamaan voidaan käyttää karttaa Jyväskylän keskusta-alueesta, katso Kuva 7 sivulta 27. Ajatellaan esimerkinomaisesti tilannetta, jossa Citymarket Keljonkeskusten Neste Oililla (Kylmälahdentie 6) bensiinin 95E hinta edellisenä päivänä oli 1.599 €/l, lähin asema ABC, Prisma Keljonkeskus (Kylmälahdentie 8) myy tänä päivänä hintaan 1.649 €/l. Oletetaan vielä, että jos Suomen keskihinta edellisenä päivänä oli 1.620 €/l ja tänään noussut 1.65 euron litrahintaan, niin mallin mukaan ennuste Neste Oilin uudeksi hinnaksi on

$0.00 + 0.66 \times (1.599 - 1.62) + 0.13 \times (1.649 - 1.62) + 1.65 = 1.640$ (€/l, tarkoilla arvoilla laskettuna). Luottamusväli 95 %:n merkitsevyytasolla tälle ennusteelle on [1.638, 1.642]. Neste Oilin hinta oli siis eilen ollut 2.1 senttiä alle keskiarvon. Tämän jälkeen keskihinnat nousivat kolmella sentillä, ja kilpailijalla päivän hinta oli käytännössä keskimääräinen hinta. Neste Oil saa täten uudeksi hinnakseen 1.0 senttiä keskiarvon alapuolella olevan hinnan. Puretaan seuraavaksi auki vielä muuten sama tilanne, mutta kilpailijan Prisman ABC:n hinta olisikin ollut kyseisenä päivänä sama kuin Nesteen sen hetkinen hinta 1.599 €/l. Tällöin mallin mukaan hinnaksi tulisi 1.633 €/l. Sama voitaisiin toistaa muillekin kartan kilpailijapareille.

Poraudutaan mallin käyttöön vielä Jyväskylän keskustan Hannikaisenkadun asemilla 'ABC, Hannikaisenkatu 6' ja 'Shell, SOK Veturi Hannikaisenkatu 20'. Haetaan mallin avulla Shellille dieselin uusi hinta, kun edellisenä päivänä aseman hinta oli 1.569 €/l, Suomen keskihinta oli 1.50 €/l ja kyseisenä päivänä keskihinta saman verran. Yleistasossa ei siis tapahtunut muutoksia. ABC myy kyseisenä päivänä dieseliä hintaan 1.539 €/l. Hintaeroa olisi tällöin kolme senttiä per litra ilman hintamuutosta. Malli antaa dieselin uudeksi hinnaksi 1.550 €/l. Luottamusväli tälle on [1.549, 1.551] €/l. Jos nämä edelliset luvut olisivat olleet muuten samat, mutta kilpailija ABC:n nykyhintana olisi ollut sama kuin Shellin edellinen eli 1.569 €/l, niin uusi hinta mal-

lin mukaan olisi 1.554 €/l. Edelleen, jos tästä keskihinta olisi noussut yhden sentin arvoon 1.51 €/l, niin hintamalli antaisi Shellin dieselin hinnaksi 1.563 €/l.

Bensiinilaadulla 95E ja dieselillä aseman uusi hinta on karkeasti yhteydessä kahteen kolmasosaan edellisestä hinnasta. Toistakymmentä prosenttia jää paikalliselle kilpailulle. Polttoainelaaduilla 98E oma hintahistoria saa suuremman painoarvon. Johtopäätöksenä tuloksista voi sanoa, että polttoaineiden vähittäiskaupassa kuitenkin on kilpailua keskimäärin. Toki syrjäseuduilla kilpailu ei ole niin kovaa eivätkä asemat ole välttämättä edes lähellä toisiaan. Aineistossa suurin osa asemista on Suomen suurimmista maakuntakeskuksista, joissa asemat ovat sijoittautuneet toistensa läheisyyteen. Tulosten perusteella asemat seuraavat toistensa hinnoittelua ja asettavat hintansa osittain näiden perusteella. Tämä on kuluttajille suotuisaa kilpailun vetäessä hintoja alaspäin. Osittain asemien vahvaa omaa hinnoittelutyyliä voi selittää esimerkiksi asemaketjujen bonuskorttijärjestelmät, liikenteellisesti hyvät sijainnit, ansaintalogiikan perustuminen muuhun oheismyyntiin, tankkaajien preferenssit, ketju-uskollisuus ja sattuma.



Kuva 7: Jyväskylän keskusta-alueen polttoaineiden markkina-alue. Lähde: Google-Maps.

4.2 Tutkimuskysymys 2: Miten mallintaa polttoaineiden hin- nankehitystä?

4.2.1 Lähtökohdat

Tässä aliluvussa tutkitaan, miten polttoaineiden hintoja voidaan yleisesti ottaen en-
nustaa muiden tekijöiden avulla. Hypoteesina on myös, että huoltoasemat hyötyisi-
vät raakaöljyn laskevasta hinnasta ja jättäisivät polttoaineiden hinnat ylihintaisik-
si joksikin aikaa Näin ei pitäisi käydä kilpailullisilla markkinoilla [18]. Luovutaan
nyt paikallisesta asemakohtaisesta tarkastelusta ja siirrytään polttoaine.net-aineiston
viikkohintojen aikasarjaan. Havaintoja on nyt 141 kappaletta kustakin polttoaine-
laadusta. Tässä käsitellään koko Suomen keskihintoja, joten tutkimuskysymyksessä
1 ollut seutukuntarajoite on poistettu. Tutkimuskohteena on selvittää polttoainei-
den hintojen yhteys raakaöljyn maailmanmarkkinahintaan, vuodenaikaan, verotuk-
seen ja 95-oktaanisen bensiinin etanolipitoisuuteen. Raakaöljy on bensiinin ja diese-
lin pääraaka-aine, joten hintayhteys on luonnollinen. Vuodenajan yhteys hintoihin on
myös lähtökohtaisesti luonteva selvittää, koska liikenteen volyymissa on kuukausista
johtuvaa kausivaihtelua [15] ja dieselistä myydään kesä- ja talvilaatuja vuodena-
jasta riippuen. Polttoaineiden hinnat sisältävät arvonlisäveron ja valmisteveron: tutkitaan
miten muutokset siirtyivät hintoihin. Polttoaineisiin sovelletut verot on esitetty Tau-
lukossa 1 sivulla 5. Selvitetään myös, miten vuonna polttoainelaadun 95E etanoli-
pitoisuuden nosto viidestä prosentista kymmeneen prosenttiin näkyi sekä 95E:n että
98E:n ja dieselin hinnoissa. Käytetään hintojen ennustamisessa vielä edellisen kuu-
kauden inflaatiota, joka tehdään kuluttajahintaindeksin luonnollisen logaritmin avulla
[14]. Koska indeksi julkaistaan kuukausittain, on viikkotason luvut johdettu lineaar-
isella interpoloinnilla. Inflaatio on selittävänä muuttuja mukana kuvaamassa yleis-
tä hintakehitystä kaikista hyödykkeistä. Inflaatiota on viivästetty kuukaudella, kos-
ka polttoaineiden hinnat osittain vaikuttavat inflaatioon eli regressiossa syy-seuraus
meni myös väärään suuntaan. Viivästämällä tätä pyritään korjaamaan. Inflaation
on kuitenkin hyvä olla mallissa mukana, koska muuten ei voida luotettavasti sanoa,
johtuuko polttoaineiden hinnannousut yleisestä hintojen noususta vai raaka-öljystä.

Koska selitettävänä on nyt polttoaineiden hinnat aikasarjana, niin peräkkäiset viik-
kohinnat korreloivat. Autokorrelaation vuoksi käytetään mallina ja estimointimene-
telmänä yleistettyä pienimmän neliösumman menetelmää (Kappale 3.6). Tutkimus-
kysymyksen ratkaisuun kokeiltiin PNS-estimointia, ja jäännöksissä esiintyi voimakas-
ta autokorrelaatiota. Tämän mallinnusvaiheen PNS-estimointi ja tarkempi tarkastelu
sivuutetaan. Todetaan vain, että perinteisen PNS-estimoinnin tulokset poikkeavat ai-
ka paljon korjatuista kertoimista. Mutta koska regression oletukset eivät täyty, niin
johtopäätöksiä ei kannata tehdä tästä sen enempää.

4.2.2 Mallinnus

Tutkitaan aluksi, onko polttoainehinnoissa epäsymmetristä yhteyttä raakaöljyn hin-
tamuuutoksiin [2], [3], [5]. Rakennetaan jokaiselle kolmelle polttoainelaadulle regres-

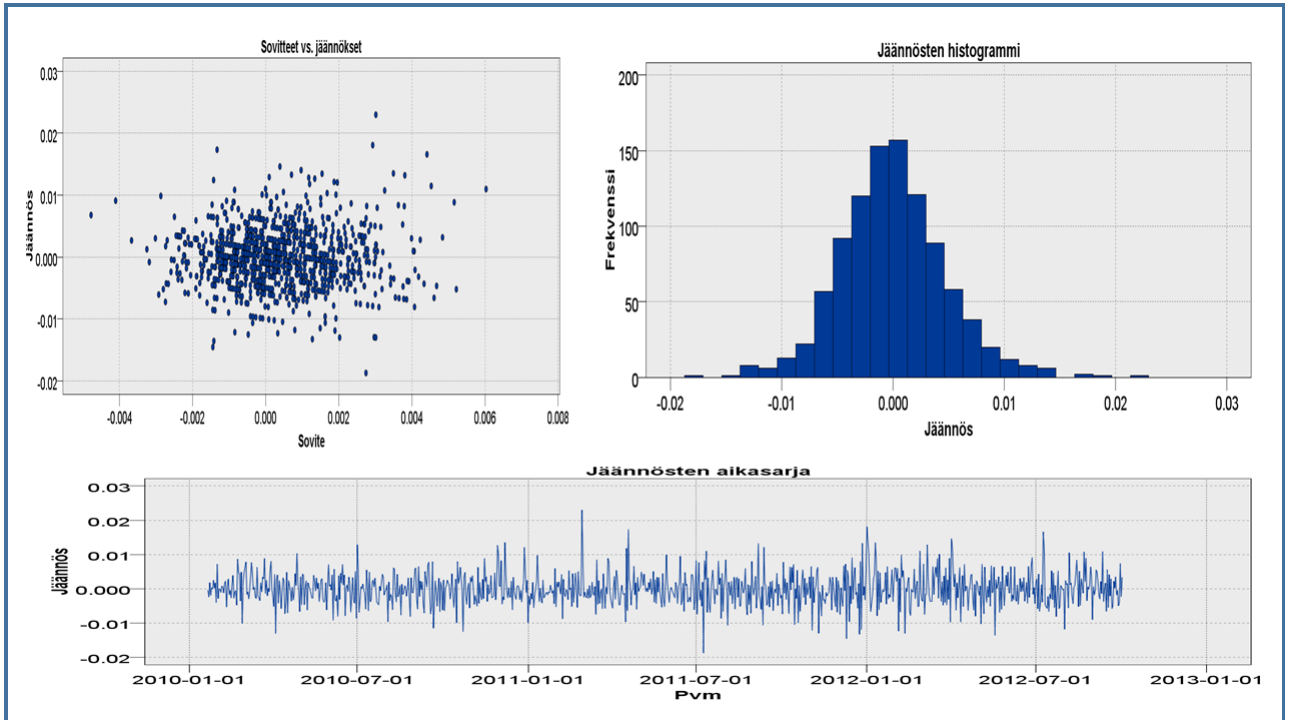
siomalli, jossa polttoaineen hinnan muutosta selitetään öljyn hinnan nousuilla ja las-
kuilla 21 edellisen päivän ajalta:

$$(4.5) \quad (y_{it} - y_{i(t-1)}) = \beta_{0,i} + \sum_{r=1}^{21} \left(\beta_{i,r}^+ \text{Brent}_{i,(t-r)}^+ + \beta_{i,r}^- \text{Brent}_{i,(t-r)}^- \right) + \epsilon_t,$$

missä y_{it} on polttoainelaadun i päivän keskihinta päivänä t ja $\text{Brent}_{i,t-r}^+ = \max\{(\text{Brent}_t - \text{Brent}_{t-r}), 0\}$ ja $\text{Brent}_{i,t-r}^- = \min\{(\text{Brent}_t - \text{Brent}_{t-r}), 0\}$. Hypoteesina on, että hin-
nat reagoivat öljyn hinnanmuutoksiin symmetrisesti. Tällöin nolla- ja vastahypoteesit
voidaan muotoilla seuraavasti:

$$H_0 : \sum_{r=1}^{21} \beta_{i,r}^+ = \sum_{r=1}^{21} \beta_{i,r}^- , \quad H_1 : \sum_{r=1}^{21} \beta_{i,r}^+ \neq \sum_{r=1}^{21} \beta_{i,r}^- .$$

F-testien mukaan kaikkien polttoainelaatujen kohdalla viitteitä epäsymmetriasta ei
ole nähtävissä. Nollahypoteeseja ei siis hylätä, p -arvot näissä olivat kaikissa yli 0.30.
Mallin 4.5 diagnostiikkakuviot on esitetty Kuvassa 8 polttoainelaadulle 95E. Tästä
nähdään, että regressiomallin oletukset ovat kunnossa, joten symmetria-analyysi
voidaan todentaa luotettavaksi.



Kuva 8: Mallin 4.5 diagnostiikkakuviot bensiinilaadulle 95E: sovitteet vs. jäännökset
(vas. ylh.), jäännösten histogrammi (oik. ylh.) ja jäännösten aikasarja (alh.).

Koska viitteitä epäsymmetriasta ei ollut, rakennetaan mallit polttoainelaaduttain viikkohinnoille ilman epäsymmetriatermejä:

(4.6)

$$\begin{aligned}
y_{it} &= \beta_{0,i} + \beta_{\text{BrentMA14},i} \text{BrentMA14}_t + \beta_{\text{ALV23},i} \text{ALV23}_t + \beta_{\text{v.vero2012},i} \text{v.vero2012}_t \\
&\quad + \beta_{\text{B95E10},i} \text{B95E10}_t + \beta_{\text{kevät},i} \text{kevät}_t + \beta_{\text{kesä},i} \text{kesä}_t + \beta_{\text{syksy},i} \text{syksy}_t \\
&\quad + \beta_{\text{log.infl},i} \text{log.infl}_{t-1} + \epsilon_{it}, \\
\epsilon_{it} &= \rho_{1,i} \epsilon_{i,t-1} + \rho_{2,i} \epsilon_{i,t-2} + u_{it},
\end{aligned}$$

missä $u_{it} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ kaikilla i, t ja

i	viittaa polttoainelaatuun ($i = \text{E95, 98E, DI}$);
t	viittaa viikkoon ($\text{Vko1 (1.1.2010), \dots, Vko134 (2.10.2012)}$),
y_{it}	on polttoainelaadun i Suomen keskihinta (€/l) päivänä t ,
BrentMA14_t	on Brent-raakaöljyn liukuva keskiarvo (€/barreli) periodilla $t - 14, \dots, t$,
ALV23_t	saa arvon 1, jos ALV:n osuus on 23 % hinnasta päivänä t ; muuten 0,
v.vero2012_t	saa arvon 1, jos valmistevero on vuoden 2012 mukainen päivänä t ; muuten 0,
B95E10_t	saa arvon 1, jos bensiinin etanolipitoisuus on 10 % päivänä t ; muuten 0,
kevät_t	saa arvon 1, jos kuukausi on maaliskuu-toukokuu päivänä t ; muuten 0,
kesä_t	saa arvon 1, jos kuukausi on kesä-elokuu päivänä t ; muuten 0,
syksy_t	saa arvon 1, jos kuukausi on syys-marraskuu päivänä t ; muuten 0,
log.infl_{t-1}	on kuluttajahintaindeksin (2010 = 100) luonnollinen logaritmi päivänä $t - 1$,
ϵ_t	on virhetermi.

Estimoidaan malli SAS-ohjelmiston AUTOREG-proseduurilla. Estimointituloksista käy ilmi, että indikaattori muuttujat vuodelle eivät ole tilastollisesti merkitseviä 95 %:n luottamustasolla. Jätetään nämä pois ja estimoidaan uudelleen malli

(4.7)

$$\begin{aligned}
y_{it} &= \beta_{0,i} + \beta_{\text{BrentMA14},i} \text{BrentMA14}_t + \beta_{\text{ALV23},i} \text{ALV23}_t + \beta_{\text{v.vero2012},i} \text{v.vero2012}_t \\
&\quad + \beta_{\text{B95E10},i} \text{B95E10}_t + \beta_{\text{log.infl},i} \text{log.infl}_{t-1} + \epsilon_{it}, \\
\epsilon_{it} &= \rho_{1,i} \epsilon_{i,t-1} + \rho_{2,i} \epsilon_{i,t-2} + u_{it}.
\end{aligned}$$

Tämän mallin estimointitulokset on esitetty Taulukoissa 8, 9 ja 10.

Malliin liittyvät diagnostiikkakuviot ovat Kuvissa 9 – 11. Kuvista nähdään, että malli on kaikinpuolin kelvollinen: sovitteet ja jäännökset eivät korreloi, jäännöksissä ei esiinny autokorrelaatiota, jäännökset ovat likimain normaalisti jakautuneita, ennusteet ovat loogisia ja selittävät muuttujat eivät korreloi jäännösten kanssa.

Taulukko 8: Mallin (4.7) estimoinnin tulokset polttoainelaadulle 95E.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,B95}$	-9.5	2.1	< 0.001
$\beta_{\text{BrentMA14},B95}$	0.0035	0.0006	< 0.001
$\beta_{\text{ALV23},B95}$	0.014	0.005	< 0.001
$\beta_{\text{v.vero2012},B95}$	0.037	0.009	< 0.001
$\beta_{\text{B95E10},B95}$	-0.022	0.009	< 0.001
$\beta_{\text{log.infl},B95}$	2.3	0.45	< 0.001
$\rho_{1,B95}$	1.13	0.036	< 0.001
$\rho_{2,B95}$	-0.21	0.037	0.03

Selitysaste 64 %.

Taulukko 9: Mallin (4.7) estimoinnin tulokset polttoainelaadulle 98E.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,B98}$	-9.7	2.1	< 0.001
$\beta_{\text{BrentMA14},B98}$	0.0038	0.0005	< 0.001
$\beta_{\text{ALV23},B98}$	0.014	0.005	< 0.001
$\beta_{\text{v.vero2012},B98}$	0.039	0.009	< 0.001
$\beta_{\text{B95E10},B98}$	-0.016	0.009	0.08
$\beta_{\text{log.infl},B98}$	2.4	0.46	< 0.001
$\rho_{1,B98}$	0.93	0.031	< 0.001

Selitysaste 64 %.

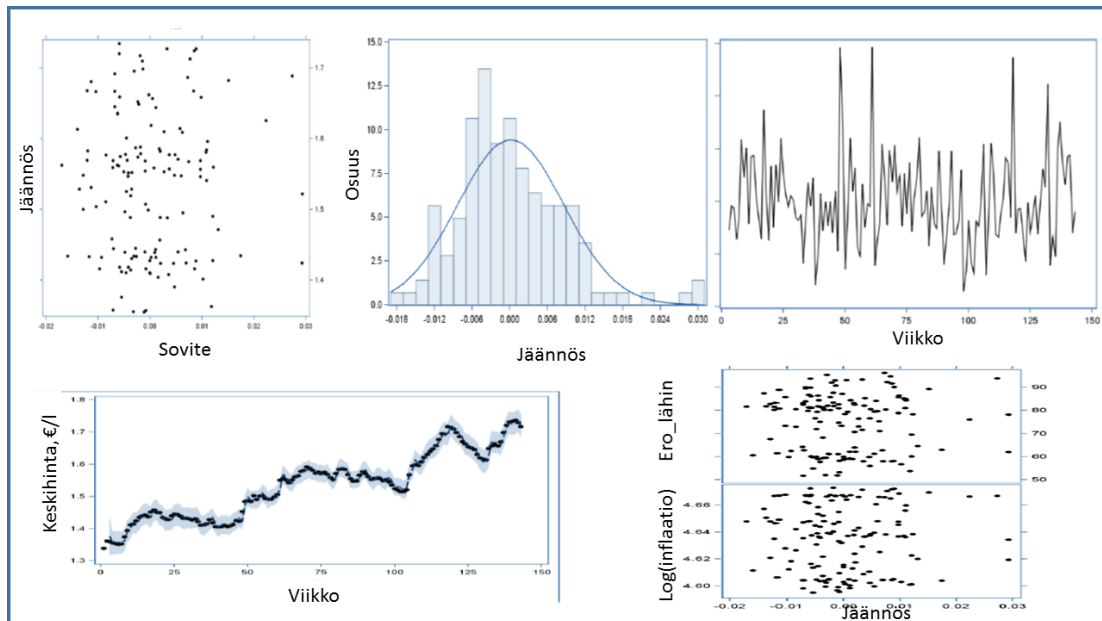
Taulukko 10: Mallin (4.7) estimoinnin tulokset polttoainelaadulle diesel.

Parametri	Estimaatti	Keskivirhe	p -arvo
$\beta_{0,DI}$	-11.8	2.6	< 0.001
$\beta_{\text{BrentMA14},DI}$	0.0037	0.0004	< 0.001
$\beta_{\text{ALV23},DI}$	0.014	0.004	< 0.001
$\beta_{\text{v.vero2012},DI}$	0.12	0.008	< 0.001
$\beta_{\text{log.infl},DI}$	2.8	0.55	< 0.001
$\rho_{1,DI}$	0.94	0.020	< 0.001

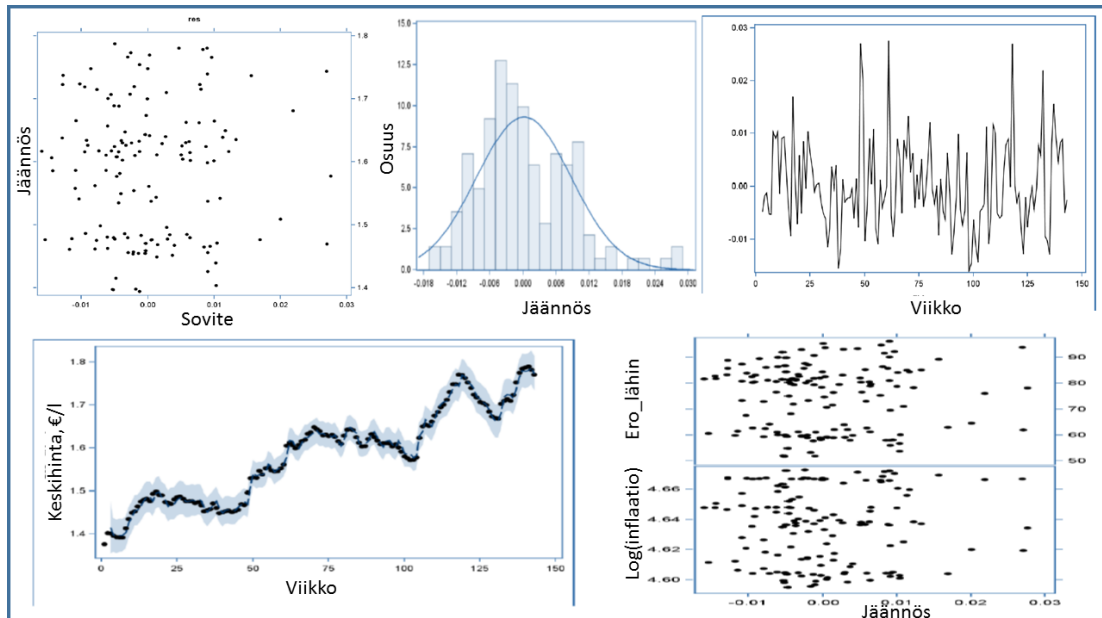
Selitysaste 77 %.

4.2.3 Tulosten tulkinta

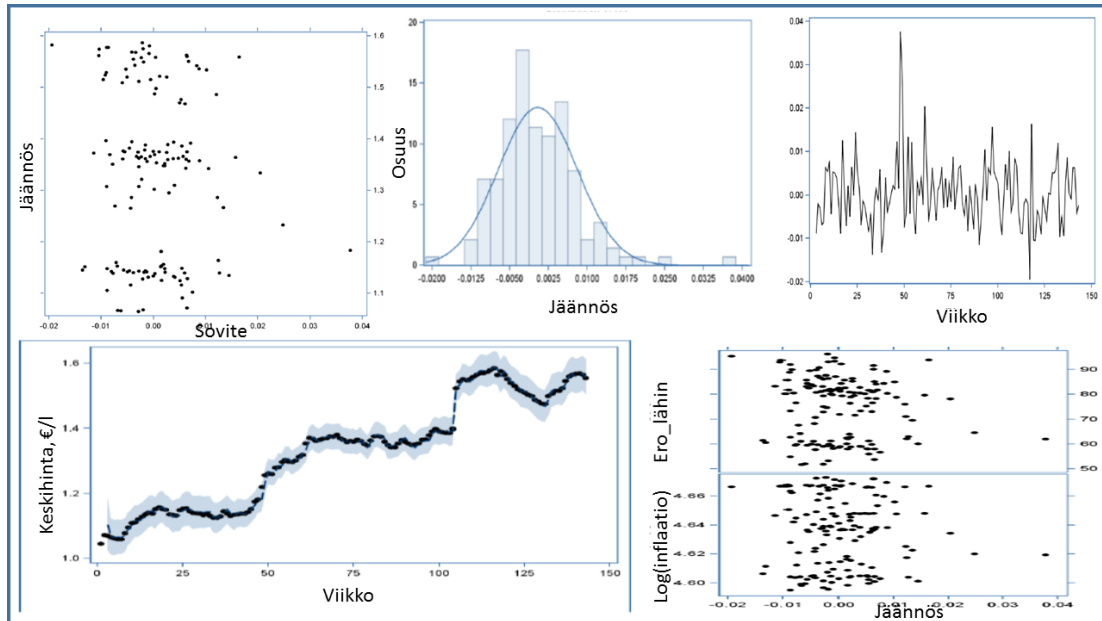
Palautetaan mieleen vielä polttoaineiden ja raakaöljyn hintojen suuruusluokka. Kuvassa 12 on piirrettynä polttoaineiden ja Brent-raakaöljyn hintojen sekä inflaation



Kuva 9: Mallin (4.7) diagnostiikkakuvioita 95E:lle: sovitteet vs. jäännökset (vas. ylh.), jäännösten histogrammi (kesk. ylh.), jäännösten aikasarja (oik. ylh.), toteutumat ja ennusteet 95 %:n luottamusvälillä (vas. alh.) sekä inflaation ja raakaöljyn arvot vs. jäännökset (oik. alh.).



Kuva 10: Mallin (4.7) diagnostiikkakuvioita 98E:lle: sovitteet vs. jäännökset (vas. ylh.), jäännösten histogrammi (kesk. ylh.), jäännösten aikasarja (oik. ylh.), toteutumat ja ennusteet 95 %:n luottamusvälillä (vas. alh.) sekä inflaation ja raakaöljyn arvot vs. jäännökset (oik. alh.).



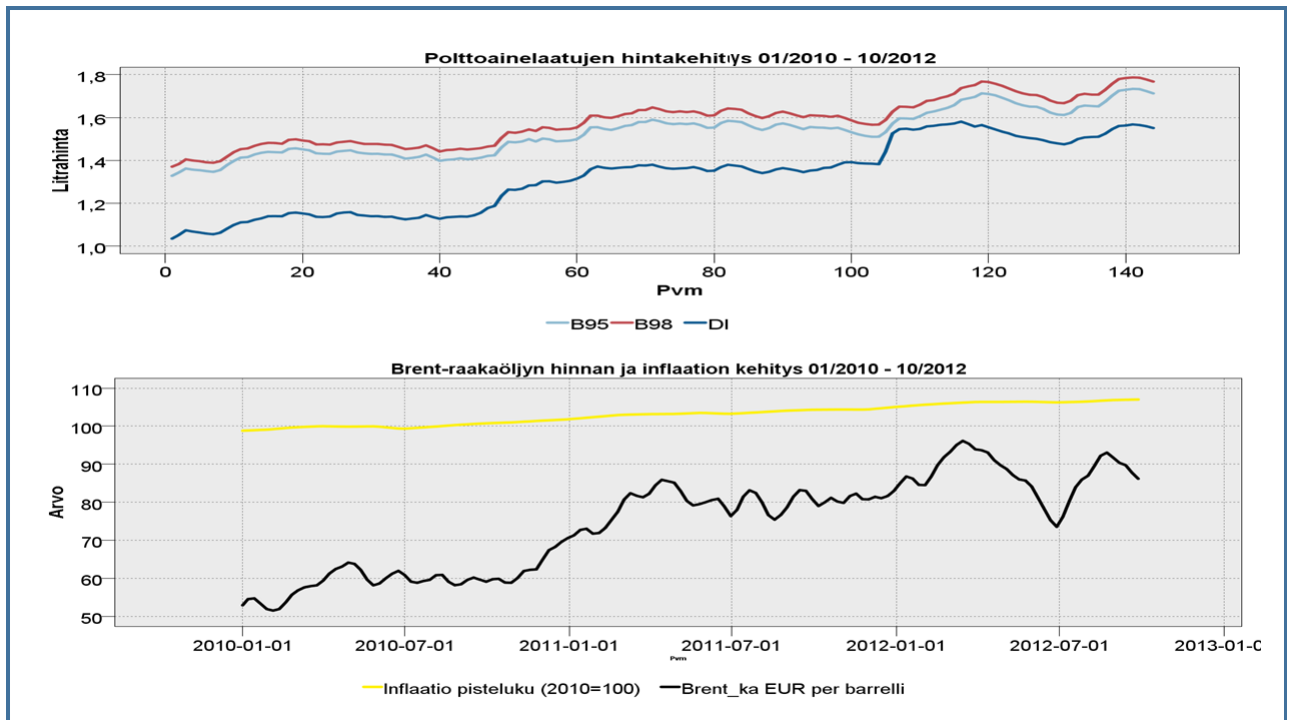
Kuva 11: Mallin (4.7) diagnostiikkakuviota dieselille: sovitteet vs. jäännökset (vas. ylh.), jäännösten histogrammi (kesk. ylh.), jäännösten aikasarja (oik. ylh.), toteutumat ja ennusteet 95 %:n luottamuvälillä (vas. alh.) sekä inflaation ja raakaöljyn arvot vs. jäännökset (oik. alh.).

kehitys aikavälillä 01/2010 - 10/2012. Taulukossa 11 sivulla 33 on esitetty näiden tunnuslukuja. Kuten estimointituloksista nähtiin, vakiotermit saavat negatiiviset arvot. Muuttujia ei ole skaalattu tai keskistetty mitenkään vaan on käytetty alkuperäisiä arvoja. Selittävien muuttujien β -kerrointen arvoihin keskistäminen ei vaikuttaisi vaan muutokset näkyisivät vakiotermissä.

Taulukko 11: Polttoaineiden ja raakaöljyn hintojen sekä inflaation kehityksen tunnuslukuja.

Muuttuja	Min	Max	Keskiarvo
95E-hinta, €/l	1.33	1.73	1.53
98E-hinta, €/l	1.37	1.79	1.58
DI-hinta, €/l	1.04	1.58	1.33
Brent-hinta, €/brl	51.54	96,14	74.48
Inflaation pisteluku	98.81	107.01	103.05
Log(inflaation pisteluku)	4.59	4.67	4.64

Mallin (4.7) tuloksista ja Taulukoista 8 - 10 nähdään, että kaikki selittävät muuttujat vuodenaikaa lukuun ottamatta ovat tilastollisesti merkitseviä 95 %:n luottamustasolla. Brent-raakaöljyn regressiokerroin bensiinilaadun 95E kohdalla 0.0035 tarkoittaa, että kun muut mallin tekijät pidetään ennallaan niin raakaöljybarrelin hinnannousu eurolla on yhteydessä 3.5 sentin hinnannousuun bensiinissä keskimäärin. Valmiste-



Kuva 12: Polttoaineiden ja Brent-raakaöljyn hintojen sekä inflaation kehitys 01/2010 - 10/2012.

vero nousi todellisuudessa 2.3 senttiä, ja mallin mukaan veron noston hintaa lisäävä vaikutus oli 3.7 senttiä. Luottamusväli (95 %) ennusteelle on [2.7 snt, 3.7 snt]. Tämän tuloksen mukaan siis polttoaineiden vähittäismyyjät nostivat 95-oktaanisen bensiinin hintaa hieman enemmän kuin valmisteveron nousu oli. Hintoja mahdollisesti myös pyöristettiin hieman ylöspäin seuraavaan .009-hintaan. Vähittäiskaupan hyötymisestä verojen noususta ei kannattane puhua, vaikka hinnannousu oli suurempaa kuin kulujen nousu: normaalihyödykkeillä kuten polttoaineet kysyntä laskee hintojen noustessa [18]. Lisäksi verojen nosto on vähittäismyynnissä haitallista. Arvonlisäverokannan muutoksella vastaava hintavaikutus on 1.4 senttiä litralta. Vastaavasti kuluttajaindeksin yhden prosentin kasvu on yhteydessä keskimäärin 2.3 sentin nousuun 95E:n hinnassa. Sivuutetaan ρ -kerrointen tulkinta, sillä niiden tehtävä oli lähinnä kontrolloida teknisesti tarkempaa estimointia. Siirtyminen suurempaan etanolipitoisuuteen on mallin mukaan yhteydessä noin kahden sentin hinnanlaskuun. Tulokset ovat käytännössä samanlaisia bensiinilaadun 98E ja dieselin osalta. Dieselin valmistevero nousi samaan aikaan 10.6 snt/l ja mallin antama ennuste vaikutukselle on 11.7 snt/l ja 95 %:n luottamusväli tälle on [10.1,13.3] snt/l. Täten ei voida siis sanoa, että asemat olisivat päässeet hyötymään tilanteesta lyhyellä aikavälillä. Todellisuudessa varmaankin veronnosto oli erittäin haitallinen myynnille, koska polttoaineiden kysyntä laskee hinnan noustessa. Dieselin mallissa on paras selitysaste, 77 %. Bensiinilaaduillekin saatiin tyydyttävät selitysasteet, 64 %.

Tarkastellaan vielä mallin antamia ennusteita annetuilla selittävien muuttujien arvoilla. Ajatellaan tilannetta, jossa kuluttujahintaindeksi on arvossa 100, myynnissä on bensiinilaatu 95E10 tarkastelujakson viimeisillä verokannoilla eli arvonlisäveroprosentti on 23 % ja valmistevero on 0.6504 €/l. Jos viimeisen kahden viikon aikana Brent-raakaöljyn hinta euroissa on ollut 75 euroa barrellilta, niin mallin mukaan ennuste hinnalle on

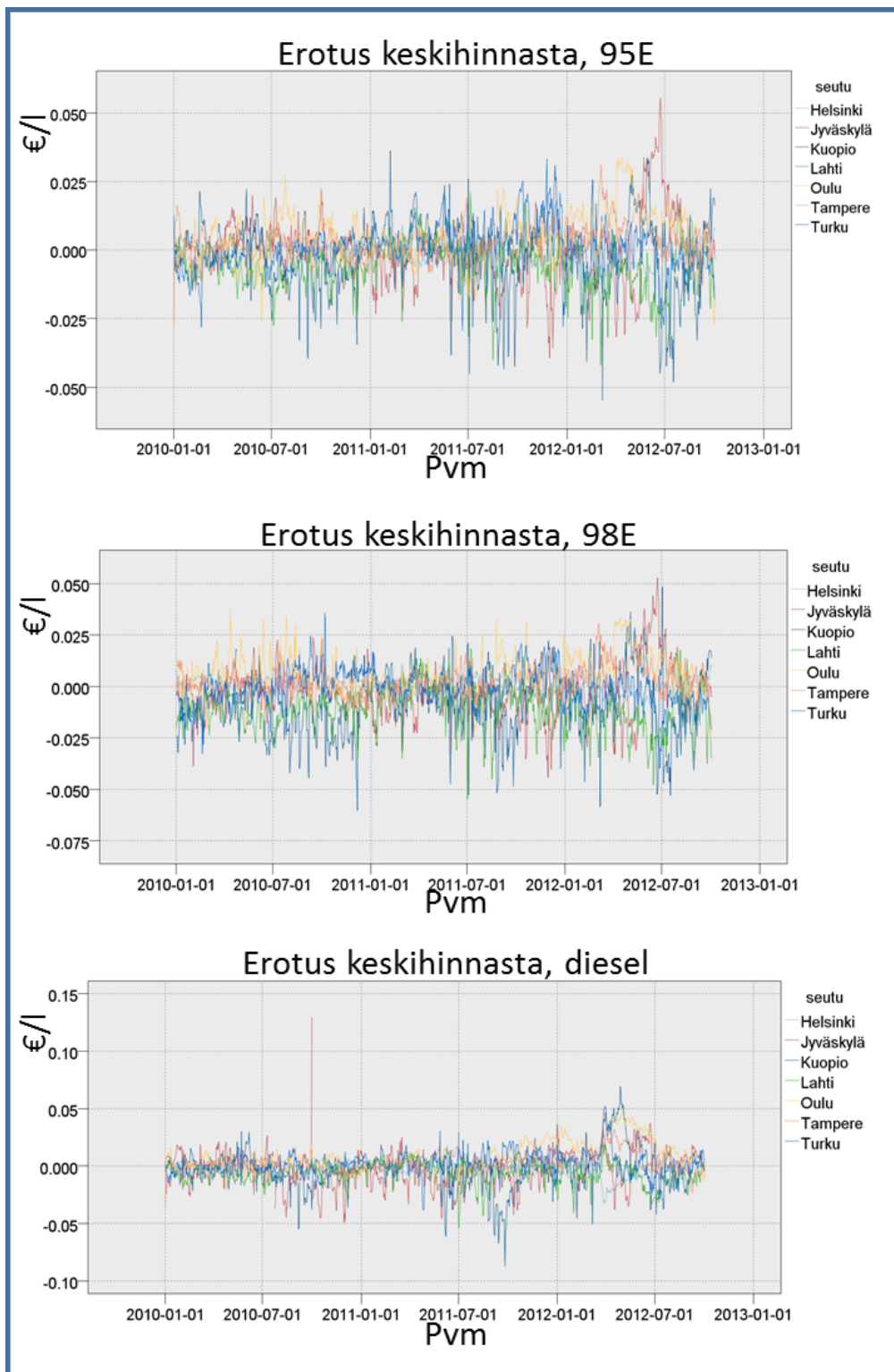
$$y_{B95,jt} = -9.5315 + 0.003541 * 75 + 0.0365 * 1 - 0.0217 * 1 + 2.3314 * \log(100) = 1.485 (\text{€/l}).$$

Jos muut tekijät pysyisivät ennallaan ja jos raakaöljyn hinta lähtisi äkillisesti nousuun siten, että kahden edeltävän viikon hinnan keskiarvo olisikin 100 euroa barrellilta, niin mallin antama hinta on 1.574 €/l. Edelleen jos tästä muut tekijät paitsi inflaatio pysyisivät muuttumattomina, niin esimerkiksi kuukauden kuluttua 100 euron barrelihinnalla ja puolen prosentin inflaatiolla edellisenä kuukautena ennuste hinnalle olisi 1.586 €/l. Tässä tapauksessa se olisi 0.7 prosentin nousu litrahinnassa. Luottamustavain leveydet edellisissä hinnoissa ovat neljä snt/l.

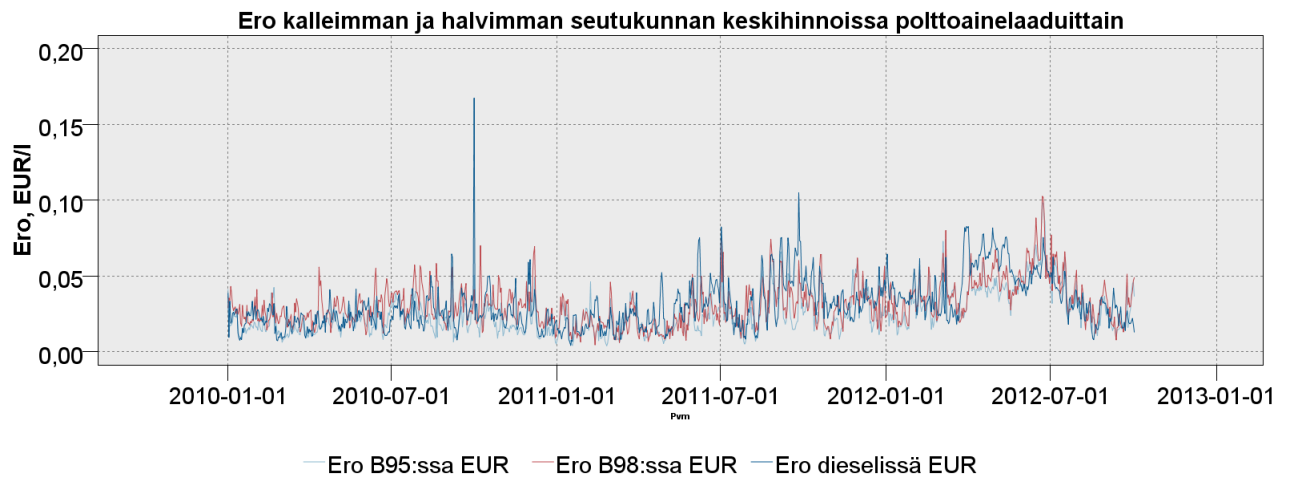
4.3 Tutkimuskysymys 3: Millaisia ovat alueelliset erot polttoaineiden hinnoissa?

Tarkastellaan, miten alueelliset erot näkyvät polttoaineiden hinnoissa tarkastelujaksona valituissa seutukunnissa. Kuvissa 13 on piirrettynä aikasarjana seutukunnittain polttoaineiden hintapoikkeama Suomen keskihinnasta. Kuvassa 14 on piirretty kalleimman seutukunnan ja halvimman seutukunnan erotus polttoainelaaduittain. Kuvista nähdään, että eri ajanhetkinä ero kalleimman ja halvimman seutukunnan keskihinnassa on muutaman sentin luokkaa - keskimäärin kolme senttiä litralta polttoainelaadusta riippumatta. Seutukuntien hintajärjestys vaihtelee ajassa, jolloin esimerkiksi Oulu löytyy sekä kalleimmalta että halvimmalta paikalta monta kertaa.

Kuvassa 15 on taulukoitu polttoainelaadun B95E10 kaksikymmentä halvinta ja kalleinta hintaa 20.3.2014 tilanteen mukaan. Halvimman ja kalleimman aseman litrahinnan erotus on jopa 24 senttiä, mikä tarkoittaa 50 litran tankkauksella jopa 12 euron eroa tankkauksen hinnalle. Myöskään noin kymmenen sentin hintaerot eivät ole ollenkaan harvinaisia. Jopa saman kunnan sisällä hintaerot voivat päästä suuriksi: esimerkiksi Espoo löytyy molempien listojen puolilta. Nurmijärvi on valloittanut kahdenkymmenen halvimman listalta puolestaan neljä sijaa, kun taas Kittilä on viidesti kalleimpien listalla. On siis löydettävissä seutukuntien sisältä klustereita, joissa hinnoittelu on yhtenäistä. Taulukossa 12 on esitetty seutukunnittain ja polttoainelaaduittain koko aineiston keskiarvot eli keskiarvot yli ajan. Jokaisen laadun kohdalla jokainen seutukunta pysyy noin sentin sisällä, joten ei löydy viitteitä, että keskimäärin jossakin otoksen seutukunnassa olisi selvästi halvempaa ja tai kalliimpaa kuin muualla. Paikallisia ilmiöitä toki esiintyy, mutta seutukuntatasolle vietäessä merkittäviä hintaeroja ei löytynyt. Taulukossa 13 on kuvattu kilpailun määrää seutukunnissa.



Kuva 13: Seutukunnan keskihinnan erotus Suomen keskihinnasta 01/2010 - 10/2012 polttoainelaaduittain: 95E (ylh.), 98E (kesk.) ja diesel (ylh.).



Kuva 14: Kalleimman ja halvimman seutukunnan hintaero polttoainelaaduittain 01/2010 - 10/2012.

Taulukko 12: Polttoainelaatujen keskihinnat koko tarkastelujaksolta (€/l).

Seutukunta	Keskihinta 95E	Keskihinta 98E	Keskihinta diesel
Helsinki	1.530	1.582	1.324
Jyväskylä	1.531	1.579	1.324
Kuopio	1.530	1.575	1.323
Lahti	1.526	1.571	1.319
Oulu	1.536	1.587	1.334
Tampere	1.535	1.585	1.328
Turku	1.531	1.580	1.326

Taulukossa on esitetty seutukunnittain eri säteillä olevien asemien lukumäärät keskimäärin. Esimerkiksi luku 1.1 Jyväskylän säteen 0.5 km kohdalla tarkoittaa, että Jyväskylän seutukunnassa asemilla on keskimäärin 1.1 kilpailijaa puolen kilometrin linnuntiesäteellä. Tässä kilpailijaksi lasketaan myös samaa ketjua olevat asemat, koska yksittäiset asemat ovat hyvin monesti yksittäisiä yrittäjiä. Taulukoissa 14 ja 15 on esitetty, kuinka monesti mikäkin seutukunta on ollut keskimäärin päivän halvin ja kallein. Useimmiten halvimpia ovat olleet Lahden ja Kuopion seudut. Kalleimpia puolestaan ovat olleet Oulun ja Jyväskylän seudut. Nämä ääritapaukset eivät kuitenkaan näyttäisi selittyvän Taulukon 13 luvuilla kilpailijamäärästä, koska yhtenäistä selkeää kaavaa ei löydy. Esimerkiksi Oulu voittaa Lahden kilpailijoiden määrässä, mutta Oulu on selkeästi ollut useammin kalliimpi kuin Lahti ja päinvastoin.

Tarkastellaan lopuksi vielä ketjukohtaisia hintaeroavaisuuksia. Tavoitteena on vastata tutkimuskysymykseen, onko jokin ketju systemaattisesti halvempi tai kalliimpi kuin toiset. Aloitetaan analyysi yksinkertaisesti tutkimalla taulukon 16 tunnuslukuja. Taulukossa on esitetty ketjuittain keskihinnat yli koko tarkastelujakson sekä monestiko

Taulukko 13: Kilpailija-asemien määrä seutukunnittain eri säteillä.

Seutukunta	$r = 0.5$ km	$r = 2$ km	$r = 5$ km	$r = 10$ km	$r = 40$ km
Helsinki	0.6	4.3	19.6	62.3	234.4
Jyväskylä	1.1	3.3	10.8	18.9	54.5
Kuopio	0.6	3.7	9.2	14.9	34.2
Lahti	0.6	2.7	7.4	12.5	50.5
Oulu	0.5	2.9	9.5	17.7	44.2
Tampere	0.5	4.2	12.7	30.1	94.3
Turku	0.6	5.4	24.5	48.4	89.1

Taulukko 14: Lukumäärät monestiko seutukunta on ollut halvin.

Seutukunta	Määrä halvimpana, kpl			Määrä halvimpana, %		
	95E	98E	DI	95E	98E	DI
Helsinki	97	14	107	10 %	1 %	11 %
Jyväskylä	194	162	269	19 %	16 %	27 %
Kuopio	271	338	227	27 %	34 %	23 %
Lahti	277	356	260	28 %	35 %	26 %
Oulu	53	26	41	5 %	3 %	4 %
Tampere	26	12	25	3 %	1 %	2 %
Turku	87	97	76	9 %	10 %	8 %

Taulukko 15: Lukumäärät monestiko seutukunta on ollut kallein.

Seutukunta	Määrä kalleimpana, kpl			Määrä kalleimpana, %		
	95E	98E	DI	95E	98E	DI
Helsinki	49	129	65	5 %	13 %	6 %
Jyväskylä	158	113	159	16 %	11 %	16 %
Kuopio	193	106	114	19 %	11 %	11 %
Lahti	40	31	31	4 %	3 %	3 %
Oulu	306	343	427	30 %	34 %	42 %
Tampere	148	147	91	15 %	15 %	9 %
Turku	111	136	118	11 %	14 %	12 %

20 Halvinta, 2 päivän sisällä päivitetty	PVM	95E10
Iitti , Neste, Kausala Matkakeidas Sitikkalantie 5	19.03.	1.569
Iisalmi , ShellExpress, Keskusta Pohjolankatu 13	20.03.	1.574
Lahti , ShellExpress, Hennala Helsingintie 120	20.03.	1.574
Nurmijärvi , Neste Oil Express, Klaukkala Klaukkalantie 59	20.03.	1.574
Nurmijärvi , ShellExpress, Klaukkala Metsäkyläntie 2	20.03.	1.574
Nurmijärvi , Teboil Express, Klaukkala Lahnuksentie 2	20.03.	1.574
Lahti , J.Kärkkäinen, Renkomäki Pasaasi 2	20.03.	1.575
Ii , J.Kärkkäinen, Sorosentie 2	20.03.	1.575
Ylivieska , J.Kärkkäinen, Kauppakeskus-Iso Ollilanojankatu 2	19.03.	1.575
Iisalmi , Neste, Eteläntie 2	20.03.	1.579
Vantaa , Neste Oil Express, Simonkallio Maitikkakuja 2	20.03.	1.579
Nurmijärvi , St1, Klaukkala Viirintie 7-9	20.03.	1.579
Kuopio , Neste Oil Express, Keskusta Tulliportinkatu 50	19.03.	1.579
Lahti , ShellExpress, Kymijärvi Siivolankatu 2	19.03.	1.589
Keuruu , Neste Oil Express, Tehtaantie 2	19.03.	1.589
Kuopio , Shell, Linja-autoasema Pohjolankatu 1 (*E99+)	19.03.	1.589
Helsinki , Neste Oil Express, Ruoholahti Salmisaarencatu 2	19.03.	1.594
Lahti , Neste Oil, Karisto Kauppiaancatu 3	19.03.	1.594
Keuruu , St1, Kangasmannilantie 20 (Re85 1.099)	19.03.	1.594
Espoo , Neste Oil Express, Lommila Lommilanrinne 1	20.03.	1.599

20 Kalleinta, 2 päivän sisällä päivitetty	PVM	95E10
Jyväskylä , Shell, SOK Veturi Hannikaisenkatu 20 (*E99+)	19.03.	1.799
Enontekiö , St1, Hetta Hetantie 2	20.03.	1.779
Nurmes , ABC, Valtatie 2-4	20.03.	1.739
Kittilä , Neste, Levi Sirkka Tunturitie 2	19.03.	1.729
Kittilä , Neste Oil (A), Lentokenttä Valtatie Lentokentän risteys	19.03.	1.729
Kittilä , Teboil, Sirkka Levinraitti 26	19.03.	1.729
Joensuu , Neste, Nollakka Lehtipojantie 1	20.03.	1.719
Outokumpu , ABC, Keskusta Keskuskatu 1	19.03.	1.719
Kittilä , ABC, S-Market, Levi Levinraitti 18	19.03.	1.719
Iloantsi , ABC, Mantsintie 2	19.03.	1.719
Kirkkonummi , Teboil, SOK Munkinmäentie 25	20.03.	1.718
Helsinki , ABC Deli, Etu-Töölö Mechelininkatu 5	20.03.	1.709
Enontekiö , Teboil (A), Hetta Ounastie 60	20.03.	1.709
Espoo , ABC Deli, Tapiola Kalevalantie 3 (E85 1.109)	19.03.	1.709
Helsinki , ABC Deli, Pitäjänmäki Vanha Viertotie 31	19.03.	1.709
Raasepori , Neste, Österby Tenholantie 2	19.03.	1.709
Kittilä , Shell, Valtatie 26 (*E99+)	19.03.	1.709
Kittilä , Neste, Valtatie 1	19.03.	1.709
Iloantsi , Neste, Kalevalantie 2	19.03.	1.709
Vihanti , Seo, Koulutie 1	19.03.	1.700

Kuva 15: Kaksikymmentä halvinta ja kalleinta asemaa 20.3.2014 (B95E10).

ketjun kalenteriviikon keskihinta on ollut halvin tai kallein. Taulukosta nousee selvästi esiin halvin ja kallein ketju. Muuten ketjujen välillä on aika tasaista. Polttoainelaudusta riippuen ABC-asetat ovat noin 2 – 2.5 senttiä litraa kohden kalliimpia kuin muut ketjut keskimäärin ovat. Vastaavasta ketju Muu on saman verran halvempi. Tämä yhdistetty luokka sisältää pienempiä kuluttajille näkyviä brändejä kuten Ysi5, J. Kärkkäinen ja MPlus. Taulukosta ilmenee myös suuri arvo Shellin kohdalla B98:n hinnassa. Tämä johtuu siitä, että vain Shellillä myytävää 98:aa korkeaktaanisempaa bensiinia tilastoidaan 98E:n tilalle. Tämä on kalliimpaa kuin 98E, mikä selittää suuren arvon ainakin pääpiirteittäin.

Taulukossa 17 on tarkastelu ketjujen hintaeroja vielä toisesta näkökulmasti parametrittömämmin. Polttoaineittain ketjut on jaettu päivämäärittäin viikon keskihinnan mukaan nousevaan järjestykseen. Joka kalenteriviikolle ketju saa siis järjestysluvun monenneksiko halvin se oli. Tämän jälkeen järjestysluvut on summattu ketjulle. Pieni arvo siis kuvastaa halpaa ketjua. Tämäkin tarkastelu osoittaa ketjun olevan halvin ja ABC:n olevan keskimäärin kallein. Tätä voi osittain selittää ABC:llä jaettavien kanta-asiakasbonusten määrä - hintaa pidetään ylempänä mutta se palautuu kuluttajille bonusten muodossa. Reaalihinta täten palautuu jonkin verran kohti muita ketjuja. Pelkkä mittaritaulun hinta - tässä tutkielmassa käytetty - ei siis aina kaikkien kuluttajien osalta kerro todellista hintaa. Osa tankkaajista varmasti tiedostaakin ilmiön ja tankkaa nimellisesti kalliimpaa polttoainetta, koska keskittämisen myötä osa hinnasta aikanaan palautuu kuluttajalle.

Taulukko 16: Ketjuittain keskihinnat yli koko tarkastelujakson sekä monestiko ketjun viikkohinta on ollut halvin tai kallein.

Ketju	Keskihinta			Lkm halvin			Lkm kallein		
	95E	98E	DI	95E	98E	DI	95E	98E	DI
ABC	1.549	1.600	1.345	0	0	0	134	5	138
Muu	1.494	1.554	1.277	139	122	135	0	0	0
Neste	1.526	1.578	1.327	0	2	2	0	0	0
SEO	1.535	1.597	1.339	0	2	0	6	1	4
Shell	1.527	1.646	1.324	3	0	4	3	137	2
St1	1.523	1.577	1.323	2	16	3	0	0	0
Teboil	1.526	1.578	1.328	0	2	0	1	1	0

Taulukko 17: Ketjuittain viikon keskihintojen järjestyslukujen summat. Suuri luku viittaa kalliiseen ketjuun.

Ketju	Järjestyslukujen summa		
	95E	98E	DI
ABC	998	862	1001
Muu	154	183	166
Neste	570	531	662
SEO	677	571	579
Shell	623	1000	620
St1	408	372	406
Teboil	602	513	598

5 Johtopäätökset

Tässä työssä tutkittiin Suomen polttoaineiden vähittäismyyntimarkkinoiden mekanismeja. Sitä varten rakennettiin tilanteisiin sopivia tilastollisia malleja, joiden avulla päästiin kvantitatiivisesti tutkimaan spatiaalisten ja temporaalisten ilmiöiden vaikutusta myyntihintoihin. Ensimmäinen kokonaisuus oli saada ymmärrystä ja edelleen luoda ennustekykyä yksittäisen aseman seuraavan päivän hinnalle. Tätä lähestyttiin kaksivaiheisesti. Ensin rakennettiin logistinen regressiomalli, jolla voidaan vastata kysymykseen "milloin asema muuttaa hintojaan". Tärkein selittävä tekijä oli hintaero naapuriaseman hintoihin. Tästä suuri osa saattoi jäädä selittämättä, koska potentiaalisia tekijöitä ovat varastotasot ja katetavoitteet. Tämän jälkeen rakennettiin lineaarinen regressiomalli uuden hinnan muodostukselle. Tässä pysähdyttiin miettimään tarkemmin paikallista kilpailua. Kilpailua löytyi jonkin verran, mutta voimakkain hinnoittelua ohjaava tekijä on kuitenkin aseman oma hinnoittelutyö. Tämä luo vahvan perustason hintatasolle, muut tekijät kilpailu mukaan lukien ohjaavat suuntaa uuden hinnan asettamiselle. Polttoainelaatua 98E muutetaan harvoiten ja se on vähiten herkkä kilpailupaineelle. Tässäkin analyysissä mallinnusvoimaa saattaisi tuoda tieto varastotasoista ja katetavoitteista.

Toinen kokonaisuus oli ennustaa yleistä polttoaineiden hintatasoa. Tässä pyrittiin tuomaan esiin globaaleja tekijöitä, joihin yksittäisillä asemilla ei ole oikeastaan edes mahdollisuutta vaikuttaa. Autokorrelaatiota korjaavalla lineaarisella regressiomallilla löydettiin selkeä yhteys raakaöljyn maailmanmarkkinahinnan ja polttoaineen loppuhinnan välille. Toinen hintaan vaikuttava tekijä on yleinen inflaatio: kun hinnat nousevat muutenkin niin silloin se näkyy myös polttoaineiden hinnoissa. Osiossa tutkittiin myös verotuksen muutosten yhteyttä hintoihin. Tuloksena oli, että verot siirtyvät käytännössä suoraan loppuhintaan sellaisenaan - ei enempää eikä vähempää. Raakaöljyn hintaheilahdusten siirtymistä loppuhintoihin analysoitiin myös. Hypoteesina oli, että vaikka öljyn hinta laskisi niin asemat saattaisivat jättää hinnat joksikin aikaa suhteellisen korkealle. Evidenssiä tästä ei kuitenkaan löydetty, mikä voi johtua tarpeeksi kilpailullisista polttoainemarkkinoista.

Viimeisessä mallinnusosiossa tarkasteltiin seitsemän seutukunnan välisiä hintaeroja. Tässä ei suuria eroavaisuuksia löydetty - Lahti on toista senttiä halvempi kuin Oulun alue. Kilpailun voimakkuudella eli asemien määrällä eri säteillä ei näyttäisi olevan tähän selvää yhteyttä. Tämän jälkeen tarkasteltiin vielä ketjukohtaisia eroja. Tässä löytyi molemmista ääripäistä poikkeuksellinen hinnoittelija. Kallein ketju on ABC ja halvin on pienien brändien koontiluokka, esimerkiksi J. Kärkkäinen Lahdessa. ABC-aseilla on käytössä S-bonus kanta-asiakasohjelma, joka sitouttaa kuluttajia keskittämään. Tämän vuoksi osa rahoista palautuu kuluttajille, ja osa saattaa tietää tankkaavansa nimellisesti kalliimpaa polttoainetta.

Polttoaineiden hinnat olivat peräisin Internet-sivuston polttoaine.net tietokannasta. Aineisto oli melko massiivinen - noin 800 000 hintahavaintoa. Aineisto oli niin sanottua avointa dataa, jota kuka tahansa voi tarkkailla. Aineistoa muokattiin paljon ja rikastettiin muista lähteistä, koska alkuperäinen aineisto sisälsi vain aseman nimen

ja päivän hinnat polttoainekohtaisesti. Aineistossa oli paljon puuttuvaa tietoa, mikä rajoittaa lopputulosten yleistämistä koskemaan täysin totuudenmukaista tilannetta. Tätä ongelmaa voidaan kuitenkin perustella sillä, että tämä on sama näkymä kenelle tahansa, joka haluaa esimerkiksi optimoida tankkausaikaa ja -paikkaa. Lisäarvoa olisi tuonut tiedot jalostamoiden ja huoltoasemien volyyymeista, varastotasoista ja kattetavoitteista. Lisäksi kokeilemisen arvoista olisi ollut käyttää tarkempaa tietoa asematyypistä. Esimerkiksi pienillä kylmäasemilla hinnoittelu on varmasti erilaista kuin ravintolan ja kaupan sisältävä moottoritien varressa olevilla asemilla. Dataa tähän ei kuitenkaan tässä työssä ollut. Jatkotarkasteluna tätä voisi tehdä. Myös asemien klusterointi ja yhteensovittaminen saatuihin tuloksiin olisi mielenkiintoinen jatkoanalyysi. Edelleen tuloksia voisi soveltaa liikennevirtoihin ja tästä voisi luoda simulointimallin.

Lähteet

- [1] Arvonlisäverolaki (1993): *30.12.1993/1501*.
- [2] Balke, N. S., Brown, S. P. A. & Yücel, M. (1998): Crude Oil and Gasoline Prices: An Asymmetric Relationship? *Federal Reserve Bank of Dallas, Economic Review*, First Quarter, pp. 2-11.
- [3] Borenstein, S., Cameron, A. C. & Gilbert, R. (1997): Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *The Quarterly Journal of Economics*, Vol. 112, pp. 305-339.
- [4] Cochrane, D. & Orcutt, G. H. (1949): Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. *Journal of the American Statistical Association*, Vol. 44, pp. 32-61.
- [5] Duffy-Deno, K. T. (1996): Retail Price Asymmetries in Local Gasoline Markets. *Energy Economics*, Vol. 18, pp. 81-92.
- [6] Zwiefelhofer, David B., findlatitudeandlongitude.com:
URL <http://www.findlatitudeandlongitude.com>. Viitattu 14.1.2013.
- [7] Greene, W. H. (1993): *Econometric Analyses*. Second Edition. Macmillan Publishing Company, New York.
- [8] Nychka, D., Furrer R. & Stephan S. (2013). Tools for spatial data. R package Fields version 6.9.1.
- [9] Krzanowski, W. J. (1998): *An Introduction to Statistical Modelling*. First Edition. Hodder Arnold, London.
- [10] Laki nestemäisten polttoaineiden valmisteverosta (1994): *29.12.1994/1472*.
- [11] Webmoon Oy, polttoaine.net: URL <http://www.polttoaine.net>. Viitattu 11.2.2013.
- [12] SAS Institute Inc., *SAS/STAT*® *Version 9.3 for Windows*. Cary, NC, USA.
- [13] Suomen Pankki: URL <http://www.bof.fi>. Viitattu 11.2.10.2012.
- [14] Suomen virallinen tilasto: Kuluttajahintaindeksi Tammikuu 2013, taulukko: Kulutushyödykkeiden keskihinnat [verkkojulkaisu]:
URL <http://tilastokeskus.fi/til/khi/index.html>. Tilastokeskus, Helsinki. Viitattu 14.2.2013.

- [15] Suomen virallinen tilasto: Tieliikenteen tavarankuljetukset 3. vuosineljännes 2013 [verkkajulkaisu]:
URL http://stat.fi/ti1/kttav/2013/03/kttav_2013_03_2013-12-20_tie_001_fi.html.
Tilastokeskus, Helsinki. Viitattu 2.3.2014.
- [16] Tilastokeskus (2011): *Suomen tilastollinen vuosikirja 2011*. Edita Prima Oy, Helsinki.
- [17] U.S Energy Information Administration (EIA): URL <http://www.eia.gov>.
Viitattu 20.2.2013.
- [18] Varian, H. R. (2010): *Intermediate Microeconomics: A Modern Approach*. Eighth Edition. W. W. Norton & Company, New York.
- [19] Yritys- ja yhteisötietojärjestelmä: URL <http://www.ytj.fi>. Viitattu 12.11.2012.
- [20] Öljyalan Keskusliitto ry: URL <http://www.oil.fi>. Viitattu 5.3.2013.