

JYVÄSKYLÄN YLIOPISTON KAUPPAKORKEAKOULU

Ville Tarvonen

**GOOGLLEN TRENDIT ASUNTOMARKKINOIDEN
INDIKAATTORINA**

Taloustiede
Pro Gradu -tutkielma
Ohjaaja: Juha Juntila
Syksy 2015

JYVÄSKYLÄN YLIOPISTON KAUPPAKORKEAKOULU

Tekijä: Tarvonen Ville Olavi	
Työn nimi: Googlen trendit asuntomarkkinoiden indikaattorina	
Oppiaine: Taloustiede	Työn laji: Pro gradu -tutkielma
Aika: Syksy 2015	Sivumäärä: 78
Tiivistelmä - Abstract	
<p>Tämä tutkimus selvittää, voidaanko Google-hakuaineistoa hyödyntää suomalaisten kerrostaloasuntojen hintojen ja niiden kauppalukumäärän indikaattorina. Monet tekijät viittaavat siihen, että suomalaisten Googlessa tekemillä verkkohauilla kyettäisiin selittämään muutoksia asuntomarkkinoilla. Suomessa lähes koko väestö käyttää internetiä säännöllisesti, Google on maan ylivoimaisesti käytetyin hakupalvelin ja lisäksi asuntomarkkinoiden palvelut ovat muuttuneet lähes kokonaan internetissä tarjottaviksi verkkopalveluiksi. Tutkimuksessa asuntomarkkinoita kuvaamaan muodostetaan kaksi eri Google-indeksiä, joista ensimmäinen indeksi pyrkii kuvaamaan yleisellä tasolla asuntoihin kohdistuvia hakuja ja toinen indeksi puolestaan koostuu Suomen suosituimpien asuntoportaaleiden nimistä.</p> <p>Saadut tulokset poikkeavat merkittävästi aiempien tutkimusten tuloksista. Koko otoksella Google-indeksien ei havaita merkittävästi parantavan asuntomarkkinamuuttujia kuvaavien mallien aineistoon sopivuutta eikä ennustetarkkuutta. Google-indeksien arvoista ainoastaan kontemporaalinen yleisellä tasolla asuntoihin kohdistuvia hakuja kuvaava Google-indeksi kykenee selittämään tilastollisesti merkittävästi asuntomarkkinamuuttujia, erityisesti asuntojen hintoja. Tämänkin Google-indeksin selityskyky on kuitenkin sensitiivinen makrotaloudellisten muuttujien huomioimiselle. Koko otoksella kummankaan Google-indeksin ei havaita Granger aiheuttavan asuntomarkkinamuuttujia, mutta robustisuustarkasteluiden yhteydessä tutkitulla pienemmällä otoksella Google-indeksien havaitaan Granger aiheuttavan asuntokauppojen lukumäärän nykyhetken arvoja, mikä viittaa Google-indeksien ja asuntomarkkinamuuttujien suhteen riippuvan jossain määrin valitusta otoksesta.</p>	
Asiasanat: Google-hakuaineisto, asuntomarkkinat, VAR-malli	
Säilytyspaikka: Jyväskylän yliopiston kauppa korkeakoulu	

KUVIOT

KUVIO 1 Internetin merkitys asunnon ostoprosessissa	11
KUVIO 2 Google-indeksin poimiminen <i>Google Trends</i> -tietokannasta	16
KUVIO 3 Sanan "iphone"-hakuintensiteetti Suomessa ja Yhdysvalloissa	18
KUVIO 4 Differoidut Google-indeksit ja asuntokaupat	51
KUVIO 5 Differoidut Google-indeksit ja asuntojen hinnat	51
KUVIO 6 Asuntokauppojen lukumäärä ja ennustemallien rekursiiviset yhden askeleen ennusteet.....	63
KUVIO 7 Asuntojen hinnat ja ennustemallien rekursiiviset yhden askeleen ennusteet.....	63

TAULUKOT

TAULUKKO 1 Kahtena eri otospäivänä poimittujen "asunto+asunnot"-indeksien arvot.....	19
TAULUKKO 2 Choin ja Varianin (2012) tulokset.....	23
TAULUKKO 3 Google-hakuaineisto ja asuntomarkkinat.....	33
TAULUKKO 4 ADF-testien tulokset	49
TAULUKKO 5 Differoitujen muuttujien kontemporaaliset korrelaatiokertoimet	52
TAULUKKO 6 Johansenin yhteisintegraatiotestin tulokset.....	53
TAULUKKO 7 Granger-kausaisuustestien tulokset	55
TAULUKKO 8 Asuntokauppojen lukumäärä ja kontemporaalinen Google1-indeksi	56
TAULUKKO 9 Asuntojen hinnat ja kontemporaalinen Google1-indeksi.....	57
TAULUKKO 10 VAR-mallin aineistoon sopivuus	59
TAULUKKO 11 Ennustemallit	60
TAULUKKO 12 Yhden askeleen ennusteet asuntomarkkinamuuttujille	62

SISÄLLYS

TIIVISTELMÄ

KUVIOT JA TAULUKOT

SISÄLLYS

1	JOHDANTO.....	7
2	ASUNTOMARKKINAT JA INTERNET SUOMESSA	9
3	INDIKAATTORIKIRJALLISUUS	12
3.1	Taloudelliset indikaattorit	12
3.2	Perinteiset indikaattorit ja niiden empiirinen soveltaminen.....	13
3.3	Google-hakuaineisto	15
3.4	Google-hakuindeksin rakentaminen	19
3.5	Google-hakuaineiston empiirinen soveltaminen.....	21
3.5.1	Google-hakuaineisto taloustieteellisenä indikaattorina.....	22
3.5.2	Google-hakuaineisto asuntomarkkinoiden indikaattorina.....	28
4	EMPIIRINEN MENETELMÄ	34
4.1	Yksikköjuuritestit.....	34
4.2	Rakenteellisten muutosten tunnistaminen	36
4.3	VAR-mallin teoria.....	38
4.4	Yhteisintegraation tutkiminen Johansenin menetelmällä	40
4.5	Ennustetarkkuuden tarkastelu	42
5	AINEISTO	44
5.1	Asuntomarkkinoiden aineisto	44
5.2	Google-aineisto	45
5.3	Makrotaloudellinen aineisto	47
6	TULOKSET.....	48
6.1	Aineiston kuvailu	49
6.2	Yhteisintegroituvuus.....	53
6.3	Granger-kausalisuus	54
6.4	Kontemporaalinen tarkastelu	55
6.5	Aineistoon sopivuus.....	58
6.6	Aineiston ulkopuolinen ennustekyky	60
6.7	Tulosten robustisuus	63
7	YHTEENVETO	67
	LÄHTEET	69
	LIITTEET.....	74

1 JOHDANTO

Taloustieteellinen ennustaminen on nojannut perinteisesti tilastoviranomaisten keräämiin aineistoihin. Tilastoviranomaisten julkaisemiin makrotaloudellisiin aikasarjoihin sekä haastattelujen avulla kerättyihin kyselytutkimuksiin liittyy ennustamisen näkökulmasta kuitenkin useita ongelmia. Nämä aineistot julkaistaan usein kuukausien viiveellä, ne ovat pitkälle aggregoituja kuvaten ainoastaan spesifin kohteen tai kuten kyselytutkimuksissa, ne voivat muodostua pienestä vastajien otoksesta. Uudet mikrotasoiset, suuriin otoksiin perustuvat ja reaaliaikaisemat mittarit voisivatkin merkittävästi parantaa nykyisiä ennustemalleja.

Internetin merkitys informaatiolähteenä on merkittävä. Kuluttajat etsivät informaatiota tavaroista ja palveluista, työttömät työpaikoista ja tiedemiehet tieteen teorioista. Informaatiotarpeet ovat yksilöllisiä, mutta yhteistä niille on se, että ensivaiheessa tiedon etsimiseen käytetään lähes poikkeuksetta hakukoneita. Hakukoneisiin kertyykin päivittäin miljardeja hakuja, jotka arkistoidaan ja joiden avulla on mahdollista analysoida ihmisten tarpeita ja niiden muutoksia.

Hakukoneiden keräämä aineisto tarjoaakin uudenlaisen mahdollisuuden ilmiöiden mallintamiseen ja ennustamiseen. Vuodesta 2006 alkaen maailman käytetyin hakukone, Google, on julkaissut ilmaiseksi viikkotasoisia tietoja eri hakusanojen suosiosta.¹ Myös taloustieteellisessä ennustamisessa on ymmärretty hakuaineiston mahdollisuudet. Hakuaineistoa on sovellettu jo useiden taloustieteellisten ilmiöiden ennustamiseen. Google-hakuaineiston avulla on pystytty parantamaan esimerkiksi työttömyyden (Askitas & Zimmermann 2009), automyyntien (Choi & Varian 2012), yksityisen kulutuksen (Vosen & Schmidt 2011 ja 2012) ja asuntomarkkinoiden (Wu & Brynjolfsson 2013) sekä osakkeiden kaupankäyntilukumäärien ennusteita (Da ym. 2011).

Suomessa Google-hakuaineiston mahdollisuuksia taloustieteellisessä ennustamisessa on tiettävästi tutkinut ainoastaan Tuhkuri (2014), joka havaitsee Google-hakuaineiston parantavan työttömyysasteen ennustemallin tarkkuutta. Hakuaineiston hyödyntämiselle myös muihin taloudellisiin ilmiöihin on kuitenkin selvät edellytykset.² Lähes jokaisella suomalaisella on pääsy internetiin ja vuonna 2014 internetiä käytti säännöllisesti 86 prosenttia väestöstä.³ Lisäksi Google on Suomessa käytännössä ainoa käytetty hakukone yli 90 prosentin markkinaosuudellaan.⁴

Tämä tutkimus laajentaa aiempaa kirjallisuutta selvittämällä Google-hakuaineiston mahdollisuuksia Suomen asuntomarkkinoiden indikaattorina. Tähän mennessä hakuaineistoa on sovellettu asuntomarkkinoiden indikaattorina tiettävästi ainoastaan kahden maan aineistoilla, yhdysvaltalaisella (ks. esim. Hohenstatt

¹Netmarketshare, Desktop Search Engine Market Share December 2014.

² Suomessa Google-verkkohakuja tehtiin vuonna 2013 keskimäärin 30 miljoonaa päivässä (Tuhkuri 2014).

³ Tilastokeskuksen Väestön tieto- ja viestintätekniikan käyttö-tutkimus (2014).

⁴ StatCounter Global Stats, Top 5 Search Engines in Finland from Dec 2008 to Aug 2015.

ym. 2011) sekä isobritannialaisella (ks. esim. McLaren & Shanbhogue 2011). Tulokset näiden maiden aineistoilla osoittavat hakuaineistosta muodostetun indikaattorin parantavan merkittävästi mallien aineistoon sopivuutta sekä ennustemallien tarkkuutta. Aiemmassa kirjallisuudessa on lisäksi korostettu Google-hakuaineiston soveltuvan erityisen hyvin pitkien ja paljon tutkimustyötä vaativien ostoprosessien kuvaamiseen (Wu & Brynjolfsson 2013).

Tarkat ennusteet Suomen asuntomarkkinoiden kokonaiskehityksestä ovat tärkeitä suomalaisille kotitalouksille, asuntomarkkinoilla toimiville yrityksille sekä sijoittajille. Suomalaiselle kotitaloudelle omistusasunnon hankinta on perinteisesti ollut elämän merkittävimpiä investointeja. Vuonna 2013 noin 68 prosenttia suomalaisista kotitalouksista asui omistusasunnossa ja omistusasunnot muodostivat yli 53 prosenttia kotitalouksien kokonaisvarallisuudesta ollen merkittävin yksittäinen varallisuuserä (Tilastokeskus 2015a).⁵ Markkinoiden kokonaiskehityksestä laaditut tarkat ennusteet ovat arvokkaita myös asuntomarkkinoilla toimiville yrityksille. Esimerkiksi rakennusalan yritysten tuloskehitys ja sitä kautta tulevaisuuden strategiset päätökset ovat vahvasti sidoksissa asuntomarkkinoiden yleiseen kehitykseen. Asuntomarkkinoiden tarkat ennusteet ovat tarpeellisia myös asuntoihin sijoittaville tahoille. Asunnoilla on oma merkityksensä esimerkiksi suomalaisten eläkevakuutusyhtiöiden sijoitusportfolioissa. Myös yksityisten kotitalouksien sijoituksia on kohdistunut yhä enemmän asuntomarkkinoille saataville tulleiden asuntorahastojen sekä sijoitusasuntojen kasvaneen suosion myötä.

Tämä tutkimus rakentuu kuudesta luvusta. Seuraavassa luvussa käydään läpi, kuinka internet on muokannut Suomen asuntomarkkinoita. Kolmas luku käsittelee indikaattorikirjallisuutta tarkastelemalla erikseen perinteisiä indikaattoreita ja Google-hakuaineistoa. Luvussa neljä puolestaan keskitytään empirian toteutuksen taustalla olevaan teoriaan. Viidennessä luvussa esitellään tutkimusaineisto ja kuudennessa luvussa raportoidaan sillä saadut tulokset. Viimeinen luku tiivistää tulokset ja pohtii niiden tutkimuksellista merkitystä suhteessa aikaisempaan kirjallisuuteen.

⁵ Omistusasuntojen osuus kotitalouksien kokonaisvarallisuudesta on pysynyt korkealla tasolla jo pitkään. Omistusasuntojen osuus kokonaisvarallisuudesta oli myös vuosina 2004 ja 2009 suurempi kuin 50 prosenttia (Tilastokeskus 2011).

2 ASUNTOMARKKINAT JA INTERNET SUOMESSA

Internet on jo pitkään ollut osa suomalaisten arkea. Tilastokeskuksen Väestön tieto- ja viestintätekniiikan käyttötutkimuksen (2005) mukaan jo vuonna 2005 73 prosenttia suomalaisista käytti internetiä viikoittain. Vuodesta 2005 eteenpäin internetin käyttö on vuosi vuodelta yleistynyt ja uudet toimintamuodot, kuten sosiaalinen media, ovat muokanneet tapaamme viestiä ja käsitellä informaatiota. Tilastokeskuksen Väestön tieto- ja viestintätekniiikan käyttötutkimuksen (2014a) mukaan vuonna 2014 60 prosentilla talouksista oli jo enemmän kuin yksi internetyhteys ja internetiä käytettiin yhä useammin teknologisen kehityksen myötä yleistyneillä mobiililaitteilla eli älypuhelimilla tai tablettitietokoneella. Vuoden 2014 tutkimuksen mukaan 86 prosenttia suomalaisista käytti internetiä säännöllisesti ja käyttäjien osuus kasvoi enää ainoastaan yli 55-vuotiaiden keskuudessa. Vaikka vuonna 2014 lähes koko väestö käyttikin internetiä säännöllisesti, käyttäjät eroavat yhä demografisten tekijöiden perusteella. Esimerkiksi perusasteen koulutuksen saaneista ainoastaan 68 prosenttia käytti internetiä, kun taas korkea-asteen koulutetuilla vastaava luku oli 96 prosenttia.

Internetin käytön yleistymisen myötä monet palvelut ovat siirtyneet verkkoon. Näin on käynyt myös asuntomarkkinoiden palveluille ja yhä useampi asunnon osto tai myyntiä harkitseva suomalainen etsiikin informaatiota asunnoista internetin välityksellä. Jo vuonna 2004 Tampereen yliopiston liiketaloudellisen tutkimuskeskuksen teettämässä kyselytutkimuksessa havaittiin, että tutkimukseen vastanneet äskettäin asunnon ostaneet henkilöt pitivät internetiä tärkeimpänä asuntokaupan tietolähteenä ja 81 prosenttia vastanneista oli käyttänyt internetiä tietolähteenään.⁶ TNS-Gallup teki puolestaan vuonna 2010 asunnon osto- ja myyntiprosessitutkimuksen, jossa selvitettiin touko-elokuussa vuonna 2010 asunnon ostaneiden suomalaisten käyttäytymistä prosessin eri vaiheissa. Tutkimuksessa havaittiin internetin olleen ensimmäinen informaatiolähde, jonka välityksellä henkilöt pääsivät kosketuksiin ostamansa asunnon kanssa. Kyselyyn vastanneet henkilöt vastasivat lisäksi hypoteettiseen kysymyksen asunnon myynnin toteuttamisesta. Vastauksista havaittiin, että internet olisi ollut ensisijainen media, johon he olisivat asunnon myynti-ilmoituksen laittaneet.⁷ Näiden tulosten perusteella voidaan todeta, että internetillä on suuri merkitys asuntomarkkinoiden tietolähteenä.

Eräs merkittävä tekijä, joka on todennäköisesti lisännyt suuresti internetin käyttöä asuntomarkkinoiden tietolähteenä, ovat sinne syntyneet asuntoportaalit. Asuntoportaalit ovat palvelleet suomalaisia jo 2000-luvun alusta saakka. Esimerkiksi Suomen suurin asuntoportaalit, Etuovi.com, perustettiin jo vuonna 2001. Asuntoportaalit keräävät suuren määrän asuntoihin liittyvää informaatiota samaan paikkaan tehden informaation hakemisen helpoksi sitä etsivälle. Huoneistokeskuksen vuonna 2011 teettämän kyselytutkimuksen mukaan ihmiset

⁶ Alma Median lehdistötiedote 1.9.2004.

⁷ Nämä TNS-Gallupin vuoden 2010 tutkimuksessa havaitut tulokset havaittiin myös vuoden 2007 tutkimuksessa.

käyttävätkin asunnon etsimiseen pääasiassa yleisesti tunnettuja ja luotettavia kanavia. Näitä ovat vakiintuneet asuntoportaalit (Etuovi.com, Oikotie.fi ja Jokakoti.fi), Googlen hakupalvelin sekä kiinteistövälitysyritysten omat verkkosivut. Näistä tärkeimpänä vastaajat pitivät asuntoportaaleja. TNS Gallupin (2013) uutiskirje vuoden 2013 Real Estate Online-tutkimuksen tuloksista kertoo asuntoportaaleiden suosion vain kasvavan vuosi vuodelta, kun taas kiinteistövälittäjien sekä rakennusliikkeiden omien sivujen merkitys asuntomarkkinoiden tiedonlähteenä pienenee jatkuvasti. Vaikka asunnon etsintä ja myynti-ilmoitukset ovatkin siirtyneet käytännössä täysin internetiin ja asuntoportaaleihin, toteutuu yhä suurin osa, noin 80 prosenttia, vuosittaisista asuntokaupoista kolmannen osapuolen eli kiinteistövälittäjän kautta (Kasso 2011). Asuntoportaaleiden käytön yleistymisen on itseasiassa lisännyt entisestään kiinteistövälittäjien avulla tehtyjen asuntokauppojen lukumäärää.⁸

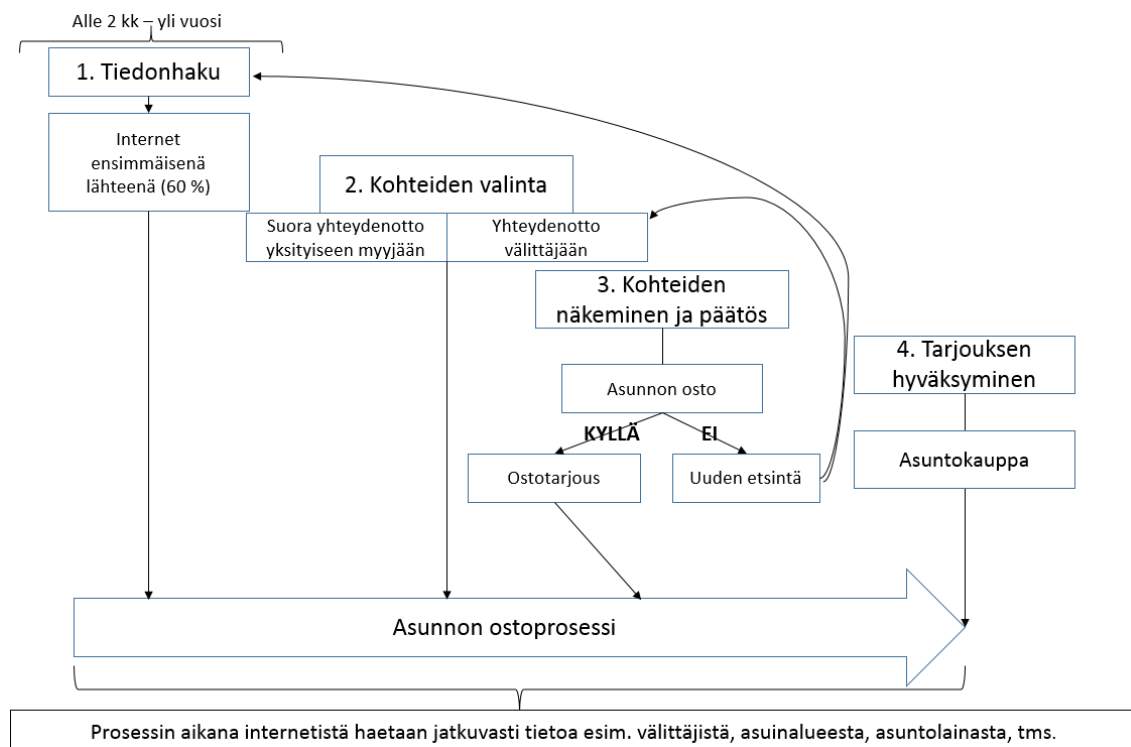
Kuviossa 1 on havainnollistettu suomalaista asunnon ostoprosessia ja internetin merkitystä sen eri vaiheissa. Kuvio pohjautuu Hohenstattin ym. (2011) kuvaamaa yhdysvaltalaiseen asunnon ostoprosessiin, jota on mukailtu Suomen asuntomarkkinoille oman subjektiivisen näkemyksen sekä TNS Gallupin (2010) asunnon osto- ja myyntiprosessitutkimuksen tietojen avulla. Kuvion 1 mukaan asunnon ostoa harkitsevan ensimmäinen askel on tiedonhaku. Ostamista suunnitteleva henkilö aloittaa asuntotarjontaan liittyvän tiedonhaun yleisimmin internetin avulla, mutta vaihtoehtoisesti käytössä voi olla muitakin lähteitä, kuten sanomalehtiä tai suusta suuhun -viestintää. Tiedonhakuprosessin kesto on yksilökohtainen, ja se voi kestää alle kaksi kuukautta, mutta pisimmillään jopa yli vuoden. Tiedonhaun jälkeen asunnon etsijä päätyy itseään miellyttävään kohteeseen tai kohteisiin, joista hän sitten kontaktoi myyjää, yleisimmin kiinteistövälittäjää. Prosessi etenee tämän jälkeen asianmukaisilla näytöillä, jonka jälkeen ostajaehdokka harkitsee, tekeekö hän tarjouksen asunnosta vai siirtyykö hän uudestaan tiedonhakuvaiheeseen etsiäkseen toista asuntoa. Takaisin tiedonhakuvaiheeseen päädytään myös, jos henkilön tekemä ostotarjous hylätään. Jos taas tehty ostotarjous miellyttää sekä myyjää että ostajaa, solmitaan asuntokauppa ja prosessi päättyy. Koko prosessin kokonaiskesto on vaikea arvioida, sillä jokainen ostoprosessi on yksilöllinen niin etsijän kuin myytävän asunnonkin ominaispiirteiden vuoksi.⁹ Huomionarvoista asunnon ostoprosessikokonaisuudessa on myös se, että internetin merkitys ei rajoitu ainoastaan tiedonhakuvaiheeseen. Internet toimii informaatiolähteenä myös silloin, kun henkilö etsii itselleen sopivaa asuntolainan tarjoajaa, tutkii asuinalueita tai asunto-osakeyhtiöitä, kerää tietoa asuntomarkkinoiden yleisestä kehityksestä tai vertailee vaikkapa välittäjien palkkioita.

Kuten edellä todettiin, internet on ollut merkittävä asuntomarkkinoiden informaatiolähde jo 2000-luvun puolivälistä ja sen merkitys on kasvanut edelleen tarjonnan keskittyessä yhä vahvemmin verkkoon. Tässä tutkimuksessa asuntomarkkinoiden kehitystä pyritään indikoimaan tunnistamalla ostoprosessin alkuvaiheen tiedonhakuvaihe, jossa internetin rooli on todistetusti merkittävä.

⁸ Asuntoportaaleissa suurin osa asuntoilmoituksista on kiinteistövälittäjien jättämiä.

⁹ Merkittävimpiä asunnon ominaispiirteitä ovat hinta, kunto ja alueellinen sijainti.

Tavoitteena on kyetä havaitsemaan henkilöiden tekemien verkkohakujen avulla asuntokysynnän vaihteluita ja ennustamaan näin asuntojen hintojen ja kauppalukumäärien kehitystä. Vaikka edellä korostettiin, että asuntoportaaleilla on itsessään suuri merkitys asuntomarkkinoilla, eivät ne ole täysin tutkimuksessa hyödynnetyn Google-hakuaineiston ulottumattomissa. Asuntoportaaleihin kohdistuvaa kiinnostusta kyetään seuraamaan Googlen raportoiman hakuaineiston avulla, jos asuntoportaalista informaatiota hakeva henkilö googlaa asuntoportaalin nimen eikä kirjoita osoitinriville kokonaista URL-osoitetta. Tutkimuksessa tehdään lähtöoletus, jonka mukaan asuntomarkkinoista kiinnostuneet henkilöt etsivät tietoa ensimmäisenä Googlesta.



KUVIO 1 Internetin merkitys asunnon ostoprosessissa

3 INDIKAATTORIKIRJALLISUUS

Tässä luvussa kuvataan aluksi, mitä indikaattorilla tarkoitetaan. Tämän jälkeen käydään läpi, kuinka perinteisiä indikaattoreita on sovellettu taloudellisten muuttujien ennustamiseen. Perinteisistä indikaattoreista siirrytään Google-hakuaineiston kuvaamiseen ja esitellään, kuinka sitä on tähän mennessä sovellettu taloudellisten ilmiöiden kuvaamisessa ja ennustamisessa.

3.1 Taloudelliset indikaattorit

Taloudellisten muuttujien kehitystä voidaan indikoida useilla erityyppisillä indikaattoreilla. Perinteisinä indikaattoreina voidaan pitää makrotaloudellisia muuttujia ja kyselytutkimuksiin pohjautuvia luottamusindikaattoreita. Nämä perinteiset indikaattorit ovat saaneet teknologisen kehityksen myötä rinnalleen myös uudenlaisia indikaattoreita. Tällaisia indikaattoreita ovat esimerkiksi Big Dataan perustuvat Google-hakuaineisto (ks. esim. Choi & Varian 2012) sekä sosiaalinen media (ks. esim. Bollen ym. 2011).

Indikaattori-sanalla tarkoitetaan muuttujaa, joka sisältää oleellista informaatiota itsensä tai toisen muuttujan nykytilasta ja tulevasta kehityksestä. Muuttuja voi olla siis myös itsensä indikaattori. Tällöin sen omien edellisten arvojen havaitaan selittävän sen nykyisiä ja tulevia arvoja. Useiden makrotaloudellisten muuttujien on havaittu olevan itsensä indikaattoreita. Esimerkki tällaisesta makrotaloudellisesta muuttujasta on Suomen asuntohintojen aikasarja (ks. esim. Englund & Ioannides 1997). Vaikka muuttuja olisikin hyvä itsensä indikaattori, muuttujan arvojen ennusteita voidaan usein parantaa lisäämällä ennustemalliin muitakin muuttujia.

Muuttujia on mahdollista hyödyntää toistensa indikaattoreina esimerkiksi silloin, kun niiden julkaisufrekvenssit eroavat (Castle ym. 2009). Esimerkiksi Tilastokeskuksen neljännesvuositasolla julkaisemaa Suomen bruttokansantuotteen kehitystä voitaisiin teoriassa indikoida kuukausittain julkaistavalla työttömyysasteella. Kuukausittain julkaistavan työttömyysasteen lopulliset arvot ovat käytössä 3-4 viikon kuluttua kuukauden päättymisen jälkeen, eli toisin sanoen työttömyysasteen uusin arvo voidaan lisätä bruttokansantuotteen ennustemalliin ennen uusimman bruttokansantuotteen arvon julkistamista.

Toinen muuttuja voi olla myös hyvä indikaattori toista muuttujaa selittävään malliin, jos indikaattorimuuttujan arvojen havaitaan edeltävän tutkittavan muuttujan arvoja. Oikarinen (2012) esimerkiksi havaitsee asuntokauppojen lukumäärää kuvaavan muuttujan arvojen edeltävän asuntohintojen nykyhetken arvoja. Tämän vuoksi asuntokauppojen lukumäärän viiveitä voitaisiinkin käyttää indikaattorina asuntohintoja ennustavassa mallissa.

3.2 Perinteiset indikaattorit ja niiden empiirinen soveltaminen

Makrotaloudellisten muuttujien hyödyntämistä itsensä ja toistensa indikaattorina käydään läpi asuntomarkkinoilla havaittujen tulosten avulla. Kyselyihin pohjautuvista luottamusindikaattoreista keskustellaan puolestaan yleisemmällä tasolla.

Jos tehokkaiden markkinoiden hypoteesi olisi voimassa asuntomarkkinoilla, ei muuttujien arvoja pitäisi pystyä ennakoimaan muuttujan omilla viiveillä eikä muiden muuttujien avulla. Tutkimuksissa on kuitenkin havaittu vahvaa todistusaineistoa tehokkaiden markkinoiden hypoteesia vastaan.¹⁰ Aikaisempi asuntomarkkinatutkimus on keskittynyt asuntomarkkinamuuttujista erityisesti asuntojen hintoihin. Useat tutkimukset ovat havainneet asuntohintojen edellisten viiveiden selittävän niiden nykyisiä arvoja (ks. esim. Case & Schiller 1989, Englund & Ioannides 1997 ja Kouwenberg & Zwinkels 2014). Asuntohintojen edelliset arvot selittävät merkittävästi muuttujan nykyisiä arvoja, mutta makrotaloudellisten indikaattoreiden merkitys asuntohintoja ennustavissa malleissa on havaittu myös suureksi. Asuntohintojen indikaattorina on onnistuneesti käytetty esimerkiksi käytettävissä olevia tuloja henkeä kohti (Case & Schiller 1990 ja 2004 sekä Oikarinen 2005), bruttokansantuotetta (Englund & Ioannides 1997), asuntolainan korkoa (Oikarinen 2005 ja Adams & Füss 2010), asuntolainakantaa (Oikarinen 2009 ja Agnello & Schucknecht 2011) sekä rakennuskustannuksia (Adams & Füss 2010). Joissain tutkimuksissa makrotaloudellisten muuttujien on havaittu selittävän lähes kokonaan asuntojen hintakehityksen. Esimerkiksi Case ja Schiller (2004) havaitsevat yhdysvaltalaisella aineistolla, että käytettävissä oleva tulo henkeä kohti kykenee selittämään lähes kokonaan asuntojen hinnoissa tapahtuneen muutoksen yli 40 tutkitussa yhdysvaltalaisessa osavaltiossa.

Tutkimustulosten perusteella tehokkaiden markkinoiden hypoteesi ei ole voimassa myöskään Suomen asuntomarkkinoilla. Englund ja Ioannides (1997) havaitsevat, että myös Suomen asuntohintakehitystä voi ennustaa hintojen omilla viiveillä sekä käyttämällä selittävinä muuttujina viiveitä bruttokansantuotteesta ja reaalisen markkinakoron kasvuasteesta. Käytettävissä olevien tulojen sekä asuntolainan koron on myös havaittu toimivan hyvänä indikaattorina Suomen asuntohinnoille (Oikarinen 2005). Muista makromuuttujista esimerkiksi asuntolainakannan on havaittu indikoivan Suomen asuntohintakehitystä (Oikarinen 2009).¹¹ Oikarinen (2009) havaitsee asuntolainakannan kehityksen olevan merkittävä indikaattori erityisesti pitkällä aikavälillä, sillä sen avulla pystytään havaitsemaan kotitalouksien likviditeettirajoituksia.

¹⁰ Tehokkaiden markkinoiden hypoteesin epäonnistumista asuntomarkkinoilla on kirjallisuudessa selitetty alhaisella likviditeetillä, suurilla transaktiokustannuksilla, rajallisella hinnoitteluinformaatiolla ja ammattimaisten sijoittajien pienemmällä määrällä sekä lyhyeksi myynnin rajoituksilla (ks. Case & Schiller 1989, Fu & Ng 2001 ja Schiller 2007).

¹¹ Oikarinen (2009) huomioi tutkimuksessaan asuntolainakannan kehityksen muodostamalla muuttujan, joka mittaa asuntolainakannan suuruutta suhteessa käytettävissä oleviin tuloihin.

Asuntolainakannan merkityksen indikaattorina havaitsee myös Vänskä (2013) pro gradu -tutkielmassaan. Hän havaitsee asuntolainakannan kasvun selittävän asuntojen hintakehitystä enemmän kuin yhdenkään muun mallissa mukana olevan makrotaloudellisen muuttujan.

Siinä missä makrotaloudellinen aikasarja perustuu numeerisiin tilastoihin, kyselyihin pohjautuva luottamusindikaattori rakentuu ilmiötä kuvaamaan suunniteltujen spesifien kysymysten vastauksista.¹² Tilastoviranomaiset julkaisevat säännöllisesti luottamusindikaattoreita useiden maiden talouksista ja niitä julkaistaan yleensä kuukausittain. Luottamusindikaattoreiden avulla pyritään kuvaamaan tulevaisuutta koskevia odotuksia, joita muut saatavilla olevat muuttujat eivät vielä sisältäisi. Usein haasteena onkin sellaisen luottamusindikaattorin löytäminen, joka tuo ennustemalliin selkeän parannuksen, kun muutkin mahdolliset indikaattorit on huomioitu.

Kyselyihin perustuvien luottamusindikaattoreiden kykyä ennakoida muutoksia taloudellisessa tilanteessa on tutkittu melko paljon. Esimerkiksi kuluttajien luottamusta kuvaavien indikaattoreiden avulla on pyritty ennakoimaan paremmin kulutuksen kehitystä. Carrol ym. (1994) havaitsevat Michiganin kuluttajaluottamusindikaattorin (*Michigan Index of Consumer Sentiment*) parantavan yhdysvaltalaisien kotitalouksien kulutuksen ennusteita. He havaitsevat kuitenkin luottamusindikaattorin merkityksen pienenevän selvästi, kun ennustemalliin lisätään makrotaloudellisia muuttujia. Wilcox (2007) lisää Carrolin ym. (1994) käyttämän Michiganin kuluttajaluottamusindikaattorin kestokulutushyödykkeiden ja palveluiden kulutusta ennustavaan malliin ja havaitsee luottamusindikaattorin huomioidun pienentävän merkittävästi mallin ennustevirheitä. Wilcox havaitsee myös luottamusindikaattorin merkityksen korostuvan erityisesti pidempää aikaväliä ennustettaessa.

Yhdysvaltojen lisäksi kuluttajaluottamusindikaattoreiden on havaittu parantavan kulutuksen ennusteita myös eurooppalaisella aineistolla. Acemoglu ja Scott (1994) havaitsevat gallup-kyselyyn perustuvan kuluttajaluottamusindikaattorin parantavan Iso-Britannian yksityistä kulutusta ennustavaa mallia, vaikka makrotaloudelliset muuttujat kontrolloitaisiinkin mallissa. Belessiotis (1996) puolestaan havaitsee ranskalaisella aineistolla kuluttajaluottamusindikaattorin kykenevän ennakoimaan maan tulevaa yksityistä kulutusta.

Vaikka useat tutkimukset ovatkin havainneet luottamusindikaattoreiden kykenevän ennakoimaan tulevaa kulutusta, on myös tutkimuksia, jotka ovat havainneet luottamusindikaattoreiden merkityksen hyvin vähäiseksi. Eräs tällainen tutkimus on Cotsomitis ja Kwan (2006). He tutkivat kuluttajaluottamusindikaattorin (*Consumer Confidence Indicator*) ja taloudellisen luottamusindikaattorin (*Economic Sentiment Indicator*) kykyä selittää kotitalouksien kulutusta yhdeksässä eri Euroopan maassa. Heidän tulostensa mukaan luottamusindikaattorit ei-

¹² Luottamusindikaattori voi myös olla yhdistelmä useista eri kyselyihin pohjautuvasta luottamusindikaattorista. Tällainen on esimerkiksi Euroopan komission luottamusindikaattori, joka on yhdistelmä teollisuuden, kuluttajien, rakennusteollisuuden sekä vähittäiskaupan luottamusindikaattoreista.

vät kykene selittämään kotitalouksien kulutuksen kehitystä kuin ainoastaan muutamassa tutkituista maista. Lisäksi luottamusindikaattoreiden selityskyky heikkenee entisestään, kun malliin lisätään makrotaloudellisia muuttujia. Ennustemallin ennustetarkkuuden he havaitsevat puolestaan heikkenevän lähes poikkeuksetta, kun ennustemalliin sisällytetään kumpi tahansa testatuista luottamusindikaattoreista.

Kulutuksen ennustamisen lisäksi luottamusindikaattoreita on sovellettu useiden muidenkin taloudellisten muuttujien kehitystä ennakoivissa malleissa. Niitä on hyödynnetty muun muassa koko talouden kehityssuunnan analysoinnissa. Esimerkiksi Taylor ja McNabb (2007) havaitsivat yritysten ja kuluttajien luottamusindikaattoreiden parantavan talouden käännekohtien ennusteita neljässä eri Euroopan maassa. Suomessa Suomen Pankki (2014) hyödyntää Cotsomitiksen ja Kwaninkin (2006) tutkimaa Euroopan komission luottamusindikaattoria bruttokansantuotteen ennustamisessa. Suomen Pankin tulokset osoittavat, että Euroopan komission luottamusindikaattori seuraa Suomen bruttokansantuotteen liikkeitä melko hyvin. Suomen Pankki kuitenkin korostaa, että Euroopan komission luottamusindeksin avulla saadaan ainoastaan karkeita ennusteita bruttokansantuotteen kehityksestä. Ennustemalliin tulee lisätä useita muitakin makrotaloudellisia muuttujia luotettavien bruttokansantuotteen ennusteiden aikaansaamiseksi.

Edellä esiteltyjen perinteisiin indikaattoreihin liittyy ennustamisen näkökulmasta useita ongelmia (ks. luku 1) ja niiden avulla kyetään selittämään muuttujien vaihtelua ainoastaan tiettyyn tarkkuuteen saakka. Usein perinteiset indikaattorit ovat myös informaatiosisällöltään päällekkäisiä. Esimerkiksi luottamusindikaattoreiden indikaatiokyvyn on usein havaittu heikkenevän merkittävästi, kun muut makrotaloudelliset muuttujat huomioidaan mallissa (ks. esim Cotsomitis & Kwan 2006). Näiden syiden vuoksi sellaisen indikaattorin avulla, joka sisältäisi uutta tai mahdollisesti reaaliaikaisempaa informaatiota perinteisiin indikaattoreihin nähden, olisi mahdollista oleellisestikin parantaa olemassa olevia ennustemalleja. Google-hakuaineistoon kertyvät miljoonat haut sisältävät suurella todennäköisyydellä sellaista henkilöiden käyttäytymisestä viestivää informaatiota, jota perinteiset indikaattorit eivät pidä sisällään.

3.3 Google-hakuaineisto

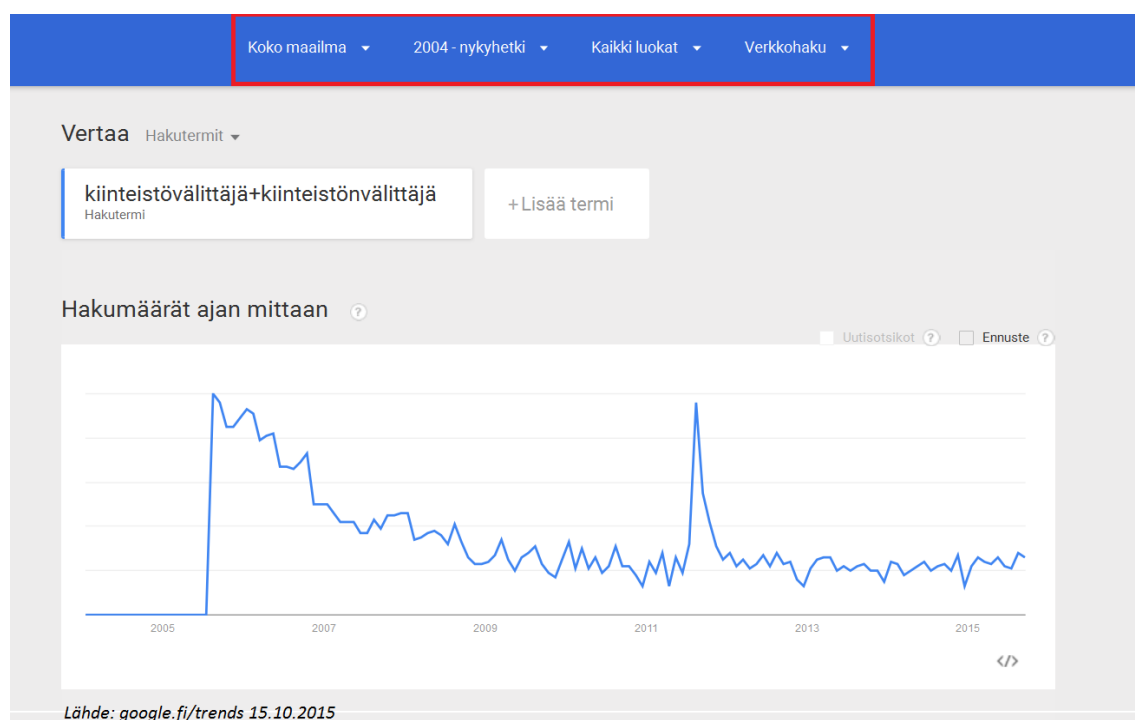
Google ylläpitää *Google Trends* -tietokantaa, joka on vapaasti kaikkien saatavilla osoitteessa google.com/trends.¹³ *Google Trends* palvelu tarjoaa tietoa Googlen hakukoneeseen kirjoitettujen sanojen yleisyydestä suhteessa kaikkiin tehtyihin hakuihin. Aineisto on saatavilla ajankohdasta 1.1.2004 eteenpäin, ja se on ladattavissa palvelusta viikkotason frekvenssillä.¹⁴ Hakuja on mahdollista tarkastella eri ajanjaksoilla, eri maantieteellisillä alueilla sekä joissain maissa myös

¹³ Suomenkielinen versio sivustosta on saatavilla osoitteessa google.fi/trends.

¹⁴ Googlen hakuaineistossa viikko alkaa aina sunnuntaista ja päättyy lauantaihin.

valmiiden luokittelukategorioiden avulla. Haetun Google-indeksin saa ladattua omalle laitteelleen .csv-muodossa.

Google Trends palvelu ei ilmoita absoluuttista hakusanalla tehtyjen verkkohakujen lukumäärää. Sen sijaan se esittää hakujen yleisyyttä, hakuintensiteettiä, kuvaavan indeksin. Palvelussa voi yhtäaikaaisesti vertailla maksimissaan viiden eri hakutermin indeksejä. Yksi hakutermi voi sisältää useita eri sanoja.¹⁵ Tämä mahdollistaa esimerkiksi sanan eri kirjoitusasujen kontrolloinnin. Jos haluttaisiin tutkia esimerkiksi kiinteistövälittäjiin liittyviä hakuja, voitaisiin eri kirjoitusasut huomioiva yhden hakutermin Google-indeksi rakentaa seuraavasti: "kiinteistövälittäjä+kiinteistönvälittäjä".¹⁶ Tällä tavoin muodostetun indeksin hakuintensiteetin laskennassa huomioidaan kaikki ne haut, joissa on mainittu toinen valituista sanoista, eli toisin sanoen hakuintensiteetin laskennassa huomioitavan verkkohaun ei tarvitse muodostua pelkästä kiinteistövälittäjä- tai kiinteistönvälittäjä-sanasta. Indeksissä sisältyy myös seuraavat kuvitteelliset esimerkkiverkkohaut: "kiinteistövälittäjä helsingissä" ja "kiinteistövälittäjä kaisa".



KUVIO 2 Google-indeksin poimiminen Google Trends-tietokannasta

Kuviossa 2 on havainnollistettu Google Trends palvelun visuaalista näkymää esittämällä hakutermistä "kiinteistövälittäjä+kiinteistönvälittäjä" muodostettu Google-indeksi. Kuvion 2 yläreunassa sijaitsevalla punaisella laatikolla on koros-

¹⁵ Yhteen hakutermiin voi yhdistää maksimissaan 25 sanaa.

¹⁶ Hakutermi huomioi molemmat sanat, kun niiden väliin asetetaan "+"-merkki. Palvelussa on mahdollista myös rajata joitain sanoja pois indeksistä. Tämä tapahtuu kirjoittamalla "-"-merkin sanojen väliin.

tettu Google-indeksin rajaushmahdollisuuksia. Maantieteelliseksi alueeksi voi valita haluamansa maan, joissain maissa jopa osavaltioita. Ajanjakson voi puolestaan vapaasti rajata vuoden 2004 toisesta viikosta eteenpäin, ja aineisto täydentyy uudella viikolla noin kahden päivän viiveellä viikon päättymisestä (Rochdi & Dietzel 2015). Hakusanoja voi lisäksi rajata Googlen valmiiksi määrittelemien luokittelukategorioiden avulla, joita on mahdollista käyttää myös itsenäisinä indekseinä.¹⁷ Valmiita kategorioita on ensimmäisellä tasolla noin 30 ja seuraavalla tasolla noin 250 (Choi & Varian 2012). Luokittelukategorioiden maantieteellisessä saatavuudessa on kuitenkin eroja. Esimerkiksi Suomen maantieteellisellä alueella näitä luokittelukategorioita ei valitettavasti ole saatavilla. Viimeisenä rajaushmahdollisuutena on käytetty hakutyyppiä. Indeksia muodostettaessa hakutyypin voi halutessaan vaihtaa verkkohauista esimerkiksi kuvahakuja koskevaksi.

Kuten edellä jo todettiin, Google tarjoaa sanan hakuintensiteettiä kuvaavan indeksin. Tämä indeksi saadaan jakamalla tietynä ajankohtana (ja maantieteellisellä alueella) tehtyjen yksittäiseen sanaan liittyvien hakujen lukumäärä saman ajankohdan (ja maantieteellisen alueen) kaikkien hakujen lukumäärällä. Tämä luku normalisoidaan jakamalla se sen ajankohdan suhdeluvulla, jona yksittäisen sanan hakujen osuus kaikista tehdyistä hauista on suurin. Tämä luku kerrotaan vielä lopuksi sadalla, jolloin saadaan välille 0-100 normalisoitu hakuintensiteetin arvo. Joidenkin hakusanojen Google-indeksin hakuintensiteettinä saattaa näkyä myös arvo nolla. Tämä johtuu Googlen hakusanoille asettamasta minimihakumäärärajoitteesta, jonka mukaan jos sanan hakujen absoluuttinen kokonaismäärä ei ylitä vähintään 50 hakua, hakuindeksin arvona näkyy arvo nolla (Choi & Varian 2011).

Kaava 1 esittää ajanhetken t hakuintensiteetin arvon laskennan aikavälillä $\{1, \dots, t, \dots, T\}$.¹⁸ S_t on tietyllä hakusanalla s tehtyjen hakujen määrä ajanhetkellä t ja H_t on kaikkien hakukoneella tehtyjen hakujen kokonaislukumäärä ajanhetkellä t . Jakajana kaavassa on valitulla aikavälillä, $\{1, \dots, t, \dots, T\}$, havaittu suhdeluvun $\frac{S}{H}$ maksimiarvo.

$$HI(S_t) = \left\{ \frac{S_t}{H_t} \right\}_{\max\left(\frac{S}{H}\right)} \times 100 \quad (1)$$

jossa

$$S = \{S_1, S_2, \dots, S_t, \dots, S_T\}$$

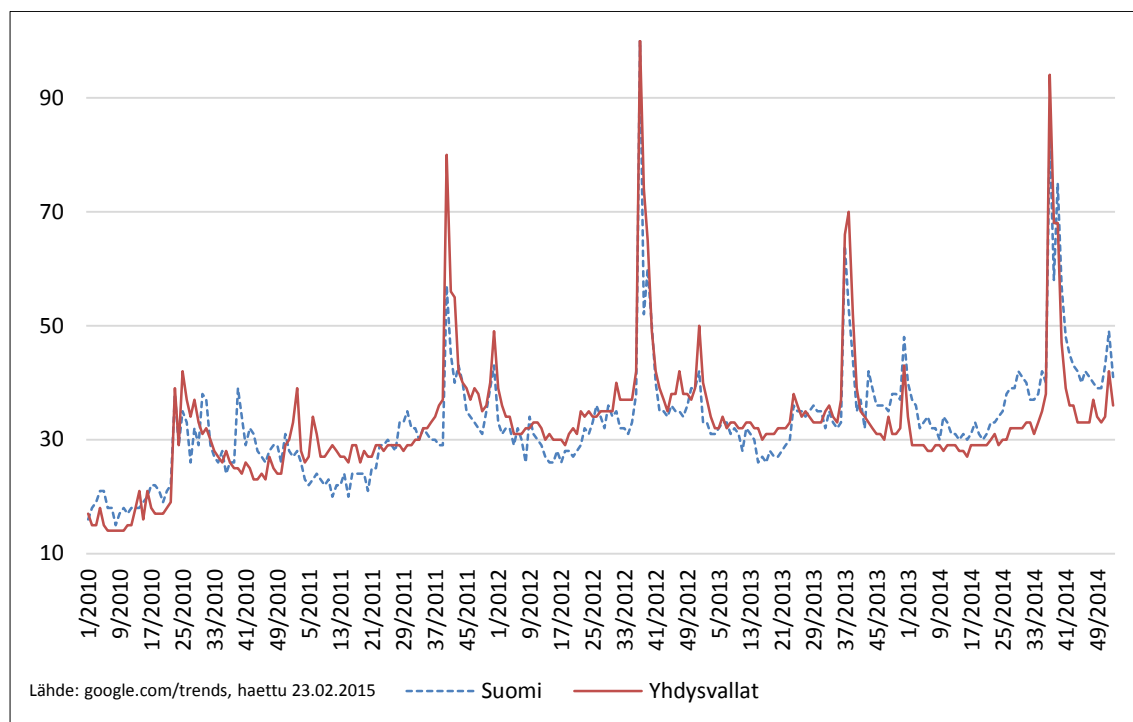
$$H = \{H_1, H_2, \dots, H_t, \dots, H_T\}$$

Koska hakuintensiteetti ei kerro hakujen kokonaismäärää, tulee hakuintensiteetti-indeksejä tulkitessa olla varovainen. Vertaamalla esimerkiksi universaalien sanan "iphone"-hakuindeksejä Yhdysvalloissa ja Suomessa ei saada tietoa, kuinka paljon hakuja todellisuudessa tehdään kyseisellä sanalla. Sen sijaan

¹⁷ Monet tutkijat ovat kirjallisuudessa soveltaneet näitä valmiita kategorioita (ks. esim. Choi & Varian 2012, Vosen & Schmidt 2011 sekä Hohenstatt ym. 2011).

¹⁸ Kaava 1 on formuloitu mukailleen Tuhkurin (2014) esitystä.

saadaan tietoa sanan suhteellisesta suosiosta näissä maissa.¹⁹ Kuviossa 4 on esitetty viikkotasoinen "iphone"-hakuindeksi Suomessa ja Yhdysvalloissa tammi-kuusta 2010 joulukuuhun 2014. Kuvioista 4 nähdään, että tarkastellulla aikavälillä sanan "iphone"-hakuindeksi (sanan suosio suhteessa kaikkiin hakuihin) on kehittynyt molemmissa maissa samankaltaisesti. Hakuindeksien kehityksessä havaitaan selvempi pitkäjaksoisempi ero ainoastaan vuoden 2010 puolivälistä vuoden 2011 puoliväliin. Molempien maiden hakuindeksit näyttäisivät myös sisältävän yhdenmukaista kausivaihtelua. Päätelmänä voidaan todeta, että "iphone"-indeksit eivät ole identtisiä, mutta sanan suhteellinen suosio on kehittynyt vertailuissa maissa hyvin samankaltaisesti.



KUVIO 3 Sanan "iphone"-hakuintensiteetti Suomessa ja Yhdysvalloissa

Internetissä tehtävien kokonaishakujen määrä kasvaa jatkuvasti. Uudet innovaatiot, kuten sosiaalinen media, ovat entisestään lisänneet internetin käyttöä ja samalla myös tehtyjen Google-hakujen määrää. Kokonaishakujen määrän jatkuva kasvu saattaa aiheuttaa laskevan trendin sanan hakuindeksiin, jollei hakusanalla tehtyjen hakujen lukumäärä kasva yleistä Google-hakujen kasvutrendiä nopeammin kuten edellisessä "iphone"-esimerkissä. Tämän internetin käyttöön liittyvän ilmiön vuoksi hakuindeksien havaitaan usein olevan

¹⁹ Esimerkiksi Barreira ym. (2013) korostaa, että maiden välisen hakujen kokonaismäärän vaihtelu on merkittävä syy Google-indeksien sisältämään kohinaan. Vaikka maiden välistä vertailua tehdessä Google-indeksit olisivat identtiset ja ennustettava muuttuja sama, tulokset voivat poiketa huomattavasti maiden hakujen kokonaismäärän vuoksi.

epästationaarisia, ja ne täytyy muuntaa stationaarisiksi robustien tulkintojen aikaansaamiseksi (Suhoy 2009).²⁰

Google-indeksit sisältävät myös otospäivästä riippuvaa vaihtelua. Google ei hyödynnä hakuindeksin laskennassa kaikkia sanalla tehtyjä hakuja. Sen sijaan se käyttää otosmenetelmää, jonka vuoksi eri päivinä poimituissa Google-indekseissä on tyypillisesti muutaman prosentin suuruinen mittausvirhe (Choi & Varian 2012). McLaren ja Shanbhogue (2011) korostavat lisäksi, että mittausvirhe on aina indeksikohtainen ja usein pienempi säännöllisesti haetuilla hakutermeillä. Taulukossa 1 on havainnollistettu ongelmaa visuaalisesti esittämällä ”asunto+asunnot”-indeksin arvot kahdeksalle viikolle kahtena eri otospäivänä. Taulukosta 1 havaitaan, että hakuindeksien arvojen erot vaihtelevat päivätasolla maksimissaan 6,3 prosenttia. Keskihajontojen ero on puolestaan valitulla kahdeksan viikon tarkastelujaksolla 8,9 prosenttia. Esitetyllä ”asunto+asunnot”-indeksillä havaitun suurehkon mittausvirheen voidaan tulkita viestivän siis siitä, että hakusanoja ei haeta kovinkaan säännöllisesti. Mittausvirhettä on kirjallisuudessa pyritty kontrolloimaan ottamalla useiden otospäivien keskiarvoja Google-indekseistä (ks. esim. McLaren & Shanbhogue 2011, Barreira ym. 2013 ja Dietzel ym. 2014).²¹

TAULUKKO 1 Kahtena eri otospäivänä poimitujen ”asunto+asunnot”-indeksien arvot

Viikko	2.3.2015	3.3.2015	%-ero
2007-12-30 - 2008-01-05	44	42	4.8 %
2008-01-06 - 2008-01-12	51	48	6.3 %
2008-01-13 - 2008-01-19	48	46	4.3 %
2008-01-20 - 2008-01-26	44	42	4.8 %
2008-01-27 - 2008-02-02	48	47	2.1 %
2008-02-03 - 2008-02-09	43	41	4.9 %
2008-02-10 - 2008-02-16	45	43	4.7 %
2008-02-17 - 2008-02-23	42	41	2.4 %
Keskihajonta	2.9	2.6	8.9 %

3.4 Google-hakuindeksin rakentaminen

Yksi Google-aineiston käytännön ongelmista on haluttua ilmiötä kuvaavan indeksin rakentaminen. Mahdollisia yksittäisiä hakusanoja on lähes rajaton määrä ja oikeiden sanojen löytäminen voi olla erittäin haastellista. Vaikka maassa olisivatkin saatavilla Googlen ylläpitämät valmiit luokittelukategoriat, saattaa

²⁰ Suhoy (2009) korostaa ongelman olevan yleinen erityisesti kulutukseen liittyvillä sanoilla.

²¹ Esimerkiksi Barreira ym. (2013) poimivat valitsemansa indeksit 14 peräkkäisenä päivänä ja havaitsivat indeksien keskihajonnan vaihtelevan 3,5 prosentista 7,6 prosenttiin.

mahdollisia ilmiötä kuvaavia kategorioitakin olla useita kymmeniä tai jopa satoja.²²

Useat tutkimukset ovat luottaneet tutkijoiden subjektiiviseen asiantuntemukseen ja niissä on käytetty sanoja tai kategorioita, jotka tutkijoiden mielestä ovat parhaiten kuvanneet tarkasteltavaa ilmiötä. Asiantuntemukseen perustuva Google-hakuindeksien valinta on ollut kirjallisuudessa selvästi yleisin tapa päätyä lopullisiin Google-indekseihin. Tutkijat ovat usein lähteneet testaamaan useita sanoja tai kategorioita ja päätyneet iteroinnin kautta mallin kannalta merkityksellisimpiin Google-indekseihin (ks. esim. McLaren & Shanbhogue 2011 ja Fondeur & Karamè 2013). Joissain tutkimuksissa on myös muodostettu useita eri Google-indeksejä kattamaan kaikki tutkijoiden mielestä oleelliset ilmiön tarkastelunäkökulmat. Esimerkiksi Askitas ja Zimmermann (2009) rakentavat useita Google-indeksejä, joista jokaisella on oma spesifi näkökulmansa suhteessa heidän tutkimaansa ilmiöön, työttömyyteen. Ensimmäinen indeksi kuvaa hakuja työvoimatoimistoihin liittyen (työttömäksi jääminen), toinen käsittää kaikki haut sanasta työttömyys, kolmas rakentuu luottamusmiestä koskevista hauista (pelko työttömyydestä) ja neljäs indeksi yhdistää työnhakua kuvaavat hakusanat. Tällainen indeksien rakennuslogiikka on Askitaksen ja Zimmermannin (2009) mukaan eräs tapa tuottaa helpommin tulkittavia Google-indeksejä.

On myös joitain tutkimuksia, joissa on sovellettu lopullisen hakuindeksin muodostamiseen tilastollisia menetelmiä. Lukuunottamatta Scottin ja Varianin (2014) tutkimusta tilastollisia menetelmiä hyödyntäneissä tutkimuksissa lähtökohtana on ollut tutkijoiden oma subjektiivinen näkemys, jonka pohjalta tilastollisilla menetelmillä käsiteltävät sanat tai kategoriat on valittu. Vosen ja Schmidt (2011 ja 2012) hyödyntävät pääkomponenttianalyysia lopullisen Google-indeksin rakentamisessa. Esimerkiksi Vosen ja Schmidt (2011) muodostavat komponentteja 56 eri Google-kategoriasta, jotka heidän mielestään parhaiten kuvaavat Yhdysvaltojen yksityistä kulutusta. Eniten yksityistä kulutusta selittävät komponentit he lisäävät lopulliseen ennustemalliinsa. Carrieré-Swallow ja Labbé (2013) puolestaan soveltavat lopullisen, automyyntejä kuvaavan, Google-indeksin muodostamiseen lineaarista pienimmän neliösumman menetelmää. He muodostavat aluksi yhdeksän erillistä Google-indeksiä Chilessä eniten myydyistä auto-merkeistä, joista he muodostavat lineaarisella pienimmän neliösumman menetelmällä yhden ennustettavaa muuttujaa parhaiten selittävän indeksin, jossa yksittäisten indeksien painot määräytyvät selityskyvyn mukaan. Scott ja Varian (2014) esittelevät puolestaan Bayesilaisen menetelmän Google-indeksien muodostamiseksi. Menetelmä pohjautuu *Google Correlate* palvelun tarjoamiin eri sanojen välisiin korrelaatioihin. Esimerkiksi työttömyyskorvaushakemusten kehitystä ennustaessaan he valitsevat 100 sanaa, joiden korrelaatio

²² Kategorioita hakuaineistona ovat käyttäneet esimerkiksi Hohenstatt ym. (2011), Choi ja Varian (2012) ja Wu ja Brynjolfsson (2013). Yksittäisistä sanoista rakennettuja indeksejä ovat käyttäneet esimerkiksi Askitas ja Zimmermann (2009), McLaren ja Shanbhogue (2011) ja Tuhkuri (2014).

työttömyyskorvaushakemusten kanssa on suurin. Näihin sanoihin he soveltavat ”spike-slap”-regressiomenetelmää, jonka avulla sanoille määritellään priorijakaumat. Näiden priorijakaumien avulla määritellään lopulliseen malliin valittavat sanat.

Julkaistujen tutkimusten perusteella Google-indeksin muodostamistavalla ei ole suurta merkitystä havaituille tuloksille. Google-indeksien on havaittu selittävän taloudellisia ilmiöitä ja parantavan niiden ennustemallien tarkkuutta olivatpa ne sitten asiantuntemuksella tai tilastollisilla menetelmillä valittuja.

3.5 Google-hakuaineiston empiirinen soveltaminen

Suurin osa Google-hakuaineistoa hyödyntävistä tutkimuksista on tehty yhdysvaltalaisella tai eurooppalaisella aineistolla. Tämä perustunee internetin käytön yleisyyteen ja Googlen merkittävään markkinaosuuteen näissä maissa.²³ Kehittyvissä maissa internet ei puolestaan ole kaikkien saatavilla. Esimerkiksi Chilessä vain 35 prosentilla (Carrieré-Swallow & Labbé 2013) ja Kiinassa 40 prosentilla (Yang ym. 2015) väestöstä on pääsy internetiin. Yhdysvalloissa vastaava luku oli puolestaan vuonna 2013 84 prosenttia.²⁴ Googllella ei myöskään ole samanlaista markkinajohtajan asemaa kehittyvissä maissa, esimerkiksi Kiinassa Googlen markkinaosuus oli vuonna 2013 ainoastaan 2,1 prosenttia (Yang ym. 2015).

Google-hakuaineistoa hyödyntäviä tutkimuksia on viime vuosina tullut yhä enemmän, sillä se tarjoaa paljon tutkimuksellisia mahdollisuuksia. Google-hakuaineisto päivittyy erittäin tiheästi ja uusin viikko on saatavilla ainoastaan muutaman päivän viiveellä viikon päättymisestä. Julkaisufrekvenssin tiheys onkin johtanut useisiin tutkimuksiin, joissa pyritään ennustamaan muuttujan nykyhetken arvoa (ks. esim. Choi & Varian 2012, Tuhkuri 2014).²⁵ Hakuaineisto on myös erittäin laaja mikrotason aineisto, josta voi eristää tietoa moniin tutkimuksellisiin ongelmiin monista eri näkökulmista. Aineiston laajuus voidaan tosin nähdä myös ongelmana, sillä joissain tapauksissa rajauksien tekeminen voi osoittautua erittäin haastelliseksi. Google-aineisto eroaa rakenteeltaan merkittävästi myös perinteisistä kyselytutkimuksista. Aineisto syntyy normaalin arkipäiväisen toiminnan sivutuotteena, minkä vuoksi perinteisten kyselytutkimusten keräämiseen liittyvät epätarkkuudet eivät välttämättä vaivaa aineistoa (Mclaren & Shanbhogue 2011). Perinteisiin kyselytutkimuksiin verrattuna Google-aineiston otos on lisäksi moninkertainen.

Kuten kaikissa aineistoissa, myös Google-hakuaineistossa on joitain ongelmia. Ongelmana voidaan pitää Google-aineiston mahdollista valikoitumisharhaa. Valikoitumisharha voi syntyä, vaikka lähes koko väestö

²³ Googlen markkinosuus Euroopassa oli yli 90 prosenttia ja Yhdysvalloissa vähintään 74 prosenttia aikavälillä 12:2008–1:2015 (StatCounter Global Stats, Top 5 Desktop, Tablet & Console Search Engines from July 2008 to Jan 2015).

²⁴ International Telecommunications Union (ITU): World Telecommunication/ICT Indicators Database.

²⁵ Kirjallisuudessa nykyhetken ennustamisesta on käytetty termiä ”nowcasting”.

käyttäisikin internetiä säännöllisesti (Askitas & Zimmermann 2015). Esimerkiksi luvussa kaksi suomalaisten internetin käyttöä tarkastellessa havaittiin, että vaikka internetin käyttöaste on maassa korkea, vaihteli se merkittävästi koulutusasteen mukaan. Tämän seurauksena Google-hakuaineiston otos saattaakin edustaa enemmän korkeasti koulutettuja kuin alhaisemman koulutustason omaavia henkilöitä. Toinen merkittävä valikoitumista aiheuttava demografinen tekijä on ikä. Tämän aiheuttamaa mahdollista valikoitumisharhaa on joissain tutkimuksissa pyritty kontrolloimaan keskittymällä pelkästään nuoriin ihmisiin, joiden keskuudessa internetin käyttö on pääsääntöisesti iäkkäämpää väestöä yleisempää (ks. esim. Fondeur & Karamè 2013). Valikoitumisongelman lisäksi Google-hakuaineistoon liittyy myös yksityisyys ja omistusoikeusongelmia, jotka ovat entisestään korostuneet mediassakin esille tulleiden tietovuotojen myötä.

Tähän mennessä Google-hakuaineistoa on sovellettu useille eri tieteenaloille. Tämän tutkimuksen kirjallisuuskatsaus pyrkii luomaan yleiskuvan Google-hakuaineiston käytöstä taloustieteellisten ilmiöiden ennustamisessa.²⁶ Taloustieteelliset tutkimukset on jaettu kahteen osaan. Aluksi esitellään taloustieteellisiä tutkimuksia yleisellä tasolla, jonka jälkeen erillisessä osiossa käydään läpi, kuinka Google-hakuaineistoa on hyödynnetty asuntomarkkinoita tutkivassa kirjallisuudessa.

3.5.1 Google-hakuaineisto taloustieteellisenä indikaattorina

Eräs tunnetuimmista Googlen hakuaineistoa käsittelevistä tutkimuksista on Choi ja Varian (2012). Se on kokoava artikkeli tutkijoiden kahdesta aiemmasta tutkimuspaperista Choi ja Varian (2009a) ja (2009b). He tutkivat artikkelissaan, voidaanko makrotaloudellisia muuttujia kuvaamaan rakennettuja yksinkertaisia malleja parantaa lisäämällä malliin kontemporaalisia Google-indeksejä. He testaavat Google-indeksien kykyä selittää ja ennustaa Yhdysvaltojen automyyntiä ja auto-osien myyntiä, viikoittaisia työttömyyskorvaushakemuksia, Hongkongin turistilukumääriä sekä Australian kuluttajaluottamusta. Choi ja Varian (2012) eivät rakenna Google-indeksejä yksittäisistä hakusanoista. Sen sijaan he hyödyntävät tuolloin käytössä ollutta *Google Insights for Search* palvelua ja sen tarjoamia kategorioita ja alakategorioita.²⁷

Taulukkoon 2 on tiivistetty Choin ja Varianin (2012) tulokset. Tuloksista havaitaan, että kontemporaalien Google-indeksin lisääminen yksinkertaiseen perusmalliin parantaa lähes poikkeuksetta mallin kykyä kuvata muuttujaa sekä ennustaa sen lyhyen aikavälin kehitystä. Poikkeuksena ovat koko ennusteperiodin aikaikkunalla saadut yhden askeleen ennusteet viikoittaisista työttömyyskorvaushakemuksista. Kontemporaalisen Google-indeksin lisääminen tähän ennustemalliin kasvattaa yhden askeleen ennusteiden absoluuttista keskivirhettä 5,95 prosenttia.

²⁶ Muilla tieteenaloilla Google-hakuaineistoa on sovellettu esimerkiksi epidemioiden ennustamiseen (Glinsberg ym. 2009 ja Malik ym. 2011), kansan mielipiteen ennakkointiin (Mellon 2014) sekä ekosysteemin tilan tutkimiseen (Proulx ym. 2014).

²⁷ *Google Insights for Search* palvelu sulautettiin vuonna 2012 osaksi *Google Trends* palvelua.

TAULUKKO 2 Choin ja Varianin (2012) tulokset

Tarkasteltu makromuuttuja	Perusmalli	Kontemporaalisen Google-indeksin lisäämisen vaikutus
Kuukausittainen autojen ja auto-osien myynnin kehitys (kausitasoittamaton)	Kausitasoitettu AR(1)-malli	Absoluuttinen keskivirhe (MAE) pienenee 10,5% koko tarkasteluperiodilla. Laman aikana, 1:2007-6:2009, MAE pienenee 21,5 %.
Viikoittaiset työttömyyskorvaushakemukset (kausitasoitettu)	AR(1)-malli	Koko ennusteperiodin aikaikkunassa MAE kasvaa yhden askeleen ennusteissa 5,95 %. Käännekohtien ennusteissa MAE pienenee 0,65 % - 21,85 %.
Hongkongin turismi (kausitasoittamaton)	Kausitasoitettu AR(1)-malli	Aineistoon sopivuus kasvaa merkittävästi, kun Google indeksi lisätään. Aineiston ulkopuolisia ennusteita ei raportoitu.
Australian kuluttajaluottamus	AR(1)-malli	Otoksen sisäinen MAE pienenee 12,7 %. Yhden askeleen ennustusteiden MAE pienenee 9,3 %.

Samoihin aikoihin, kun Choi ja Varian julkaisivat ensimmäisen kerran tutkimuksensa Google-hakuaineiston mahdollisuuksista, julkaisi Israelin keskuspankin tutkija Suhoy (2009) tutkimuksensa Googlen luokittelukategorioiden hyödyntämisestä Israelin kokonaistaloudellisen kasvuasteen ennustamisessa. Hänen tavoitteenaan on estimoida tarkkoja todennäköisyyksiä vuonna 2008 alkaneelle laskusuhdanteelle ja sitä kautta selvittää, olisiko Google-indekseillä voitu ennakoida tarkemmin laskusuhdanteen alkuajankohta. Hänen lähtöoletuksenaan on, että kategorioista muodostetut Google-indeksit pystyvät mittaamaan yleistä taloudellista toimeliaisuutta. Laskusuhdanteen indikaatioksi hän tulkitsee valittujen Google-indeksien poikkeamisen niiden pitkän aikavälin kasvu-uralta.

Kokonaistaloudellisen kasvuasteen kehitystä analysoidakseen Suhoy rakentaa Google-kategorioita sisältäviä ennustemalleja useille talouden tilaa kuvaaville muuttujille. Hän muodostaa erikseen mallin teollisuustuotannolle, vähittäiskaupalle, palveluiden liikevaihdolle, viennille sekä tuontitavaroille ja työttömyysasteelle. Malleihin valittavat Googlen kategoriat Suhoy määrittää Granger-kausaliisuustestien avulla. Google-kategorioista lopullisiin malleihin valikoituvat "human resources", "home appliances", "real estate", "travel", "food and drink" ja "beauty and personal care". Näistä kategorioista ensimmäisen, "human resources", Suhoy havaitsee indikoivan työttömyysasteen kehitystä. Kun "human resources"-kategorian hakuintensiteetti kasvaa, indikoi se työttömyysasteen nousua. Viiden muun kategorian Suhoy tulkitsee indikoivan yleistä kuluttajaluottamusta, eli toisin sa-

noen näiden kategorioiden hakuintensiteetin tulisi kasvaa hyvinä aikoina ja heikentyä huonoina. Lopuksi hän laskee nämä hakukategoriat sisältävien mallien avulla Bayesilaiset todennäköisyydet laskusuhdanteen realisoitumiselle.

Hänen tulostensa mukaan jokainen edellä mainituista Google-kategorioista auttaa ennakoimaan laskusuhdanteen alkuaikajankohtaa. Laskusuhdanteen havaitsemistodennäköisyys paranee eniten, kun malliin lisätään työttömyyttä indikoivaa ”human resources”-kategoria.²⁸ Tulostensa pohjalta Suhoy toteaa, että Google-hakuaineiston pienemmän julkaisuviiveen ja havaitun indikaatiokyvyn ansiosta laskusuhdanteiden alkuaikajankohta on mahdollista estimoida reaaliaikaisemmin verrattuna malleihin, joissa hyödynnetään ainoastaan virallisia kuluttajaluottamusmittareita.

Google-hakuaineiston mahdollisuuksia kuluttajien luottamusta indikoivana mittarina testaavat myös Vosen ja Schmidt (2011). He selvittävät tutkimuksessaan, olisivatko Googlen kategoriat parempia indikaattoreita Yhdysvaltojen yksityiselle kulutukselle kuin kyselytutkimuksiin perustuvat perinteiset kuluttajaluottamusindikaattorit, MSCI- ja CCI-indeksi.²⁹ He analysoivat kuluttajaluottamusindikaattoreita kahden perusmallin avulla. Ensimmäisenä perusmallina on yksityisen kulutuksen omia viiveitä sisältävä AR-malli ja toisessa perusmallissaan he täydentävät tätä AR-mallia yleistä taloudellista tilaa kuvaavilla makrotaloudellisilla muuttujilla.³⁰ Google-hakuaineistona he käyttävät Choin ja Varianinkin (2012) hyödyntämiä *Google Insights for Search* palvelun valmiita kategorioita, joista sopivimmat he valitsevat lopulliseen malliinsa pääkomponenttianalyysillä.

Vosenin ja Schmidtin tulosten mukaan Google-kategorioista muodostetun kuluttajaluottamusindikaattorin lisääminen AR-malliin tai makromuuttujilla täydennettyyn AR-malliin parantaa mallin aineistoon sopivuutta enemmän kuin perinteisten luottamusindikaattoreiden lisääminen. Lisäksi Google-luottamusindikaattorilla täydennetty AR-malli on tilastollisesti merkitsevästi parempi kuin MSCI- tai CCI-indeksillä täydennetty malli. Perinteisten luottamusindikaattoreiden ennustekyky tosin paranee, kun perusmalliin lisätään makrotaloudelliset muuttujat. Tällä rikkaammalla perusmallilla Google-luottamusindikaattorin sisältävä malli on tilastollisesti merkitsevästi tarkempi nykyhetken ja lähitulevaisuuden ennusteissaan kuin MSCI-indeksin sisältävä malli, mutta CCI-indeksin sisältävään malliin verrattuna tilastollisesti merkitsevää eroa ei havaita. Vosen ja Schmidt korostavat lisäksi, että Google-luottamusindikaattorin sisältämä malli kykenee lähes reaaliajassa havaitsemaan subprime-kriisin seurauksena alkaneen vuoden 2008 laman (vrt. Suhoy 2009). Perinteiset kuluttajaluottamusindikaattorit eivät kykene tähän kuin usean kuukauden viiveellä.

²⁸ Suhoy havaitsee myös Googlen ”human resources”-kategorian lisäämisen ARMA(2,2)-malliin pienentävän selvästi työttömyyden ennustevirheitä.

²⁹ MSCI-indeksi on Michiganin kuluttajaluottamusindeksi (*Michigan Consumer Confidence Index*) ja CCI-indeksi on toinen yhdysvaltalainen kuluttajaluottamusindeksi (*Conference Board Consumer Confidence Index*).

³⁰ Makrotaloudellisina muuttujina heidän mallissaan ovat: reaalitytulot, Yhdysvaltojen kolmen kuukauden valtionlainan korko ja S&P 500-osakeindeksi.

Vosen ja Schmidt (2012) tutkivat Vosenin ja Schmidtin (2011) robustisuutta toistamalla identtisen tutkimuksen Saksan yksityiselle kulutukselle. Tutkimuksessaan he vertaavat Google-kategorioista muodostettua luottamusindikaattoria Euroopan komission kuluttajaluottamusindikaattoriin (CCI-indeksi) ja Euroopan komission vähittäiskaupan luottamusindikaattoriin (RTI-indeksi). Tässäkin tutkimuksessa vertailtavina perusmalleina ovat yksinkertainen omien viiveiden AR-malli sekä makromuuttujilla (reaaliset neuvotellut palkat ja kolmen kuukauden Euribor-korko) täydennetty AR-malli.

Vosenin ja Schmidtin (2012) tulokset ovat samansuuntaisia Vosenin ja Schmidtin (2011) kanssa, tosin eivät aivan yhtä selkeätulkintaisia. Käyttäessään yksinkertaista AR-mallia perusmallinaan Vosen ja Schmidt (2011) eivät havaitse Google-luottamusindikaattorin ja perinteisten luottamusindikaattoreiden ennustetarkkuudessa tilastollisesti merkitsevää eroa. Kun perusmallina käytetään makromuuttujilla täydennettyä mallia, tuottaa Google-luottamusindikaattorin sisältävä malli tilastollisesti merkitsevästi tarkempia ennusteita kuin CCI-indeksin sisältämä malli. Tämä tulos on robusti sekä kuukausittaisille että neljännesvuosittaisille ennusteillekin. RTI-indeksin sisältävään malliin verrattuna tilastollisesti merkitsevää eroa ennustetarkkuudessa ei kuitenkaan havaita.

Carrière-Swallow ja Labbé (2013) on tiettävästi ensimmäinen kehittyville markkinoille Google-hakuaineistoa soveltanut tutkimus. He keskittyvät tutkimuksessaan selvittämään, paranevatko Chilen automyyntien nykyhetken ennusteet, kun ennustemalliin lisätään tunnetuista automerkeistä muodostettu Google-indeksi. He tutkivat Google-indeksin kykyä parantaa nykyhetken ennusteita lisäämällä sen kolmeen erilaiseen perusmalliin. Ensimmäinen perusmalli on perinteinen yhden viiveen AR-malli, toinen perusmalli on teollisuuden aktiivisuudella, IMACEC-indeksillä, täydennetty AR-malli ja kolmantena perusmallina he käyttävät ARMA(2,2)-mallia.

Carrière-Swallow ja Labbé analysoivat tulostensa pohjalta Google-indeksin merkitystä mallin aineistoon sopivuudelle, nykyhetken ennustekyvylle sekä kyvylle tunnistaa käännekohtia automyyntien aikasarjassa. Heidän tulostensa mukaan Google-indeksin lisääminen malliin parantaa tilastollisesti merkitsevästi mallin aineistoon sopivuutta. Lisäksi kulloinkin testatun perusmallin kyky ennustaa automyyntien nykyhetken arvoa paranee lähes poikkeuksetta tilastollisesti merkitsevästi, kun Google-indeksi lisätään malliin. Carrière-Swallow ja Labbé havaitsevat Google-indeksin lisäämisen myös parantavan mallin kykyä havaita käännekohtia. Esimerkiksi ARMA-malliin verrattuna tämä ero on tilastollisesti merkitsevä ja suuruudeltaan 15 prosenttia.

Barreira ym. (2013) keskittyvät tutkimuksessaan analysoimaan, vaihtelee Google-indeksejä sisältävien mallien ennustekyky maakohtaisesti. He rakentavat työttömyydelle ja automyynteille nykyhetken ennustemallit neljälle eri Euroopan maalle: Portugali, Espanja, Italia ja Ranska. Yksittäisistä sanoista muodostetut Google-indeksinsä he lisäävät perusmallina käytettyyn yksinkertaiseen ARMA-malliin. Maakohtaisia eroja he huomioivat lisäämällä malleihin "dummy"-muuttujia, joiden avulla he kontrolloivat muun muassa maiden poliittisia muutoksia,

kuten verouudistuksia.³¹ Lisäksi jos analysoitavien aikasarjojen havaittiin olevan yhteisintegroituneita, lisättiin maan ennustemalliin virheenkorjaustermi.

Barreira ym. havaitsevat Google-indeksien indikaatiokyvyn vaihtelevan maakohtaisesti. Google-indeksin lisääminen työttömyyden nykytilaa ennustavaan malliin parantaa mallin ennustetarkkuutta kolmessa neljästä testatusta maasta. Espanja on ainoa maa, jonka työttömyyden nykyhetken ennustemallia Google-indeksin lisääminen ei paranna. Automyyntien osalta heidän tuloksensa ovat päinvastaiset kuin Choin ja Varianin (2012) ja Carrière-Swallowin ja Labbén (2013). Kun Google-indeksi lisätään automyyntien ennustemalliin, Barreira ym. eivät havaitse mallien ennustetarkkuuden paranevan yhdessäkään tarkastelluista maista.³² Barreira ym. selittävät eroa automyyntien tuloksissa Google-indeksien kohinalla, jonka merkitys vaihtelee maittain. Lisäksi he korostavat, että ihmisten tavat etsiä tietoa verkosta eroavat maittain ja eivät ole ajassa vakioita. He korostavat myös, että suurien kulutustavaroiden hankinnassa ihmisten käyttäytyminen poikkeaa suuresti, mikä vaikeuttaa mallin viiveiden määrittystä. Toiset ostavat esimerkiksi auton lyhyen etsimisen jälkeen, kun taas toisilta etsintäprosessi saattaa kestää useita vuosia.

Työttömyyden indikaattorina Google-hakuaineistoa on käytetty useissa muissakin tutkimuksissa kuin jo mainituissa Choin ja Varianin (2012) ja Barreiran ym. (2013). Suomessa Tuhkuri (2014) tutkii työttömyyteen liittyvistä sanoista rakennetun Google-indeksin merkitystä työttömyyden nykyhetkeä ja lähitulevaisuutta ennustavalle mallille.³³ Perusmallinaan Tuhkuri käyttää kausitasoitettua AR-mallia.

Tulokset ovat hyvin samankaltaisia suhteessa aiempaan kirjallisuuteen. Google-indeksin lisääminen kausitasoitettuun AR-malliin pienentää mallin absoluuttista keskivirhettä ja parantaa mallin aineistoon sopivuutta. Google-indeksin sisältämä malli tuottaa 10 prosentin parannuksen ennustetarkkuuteen koko ennustetun otoksen aikaikkunassa. Tuhkuri tarkastelee nykyhetken ennusteita erikseen myös lama-aikana eli vuosina 2008-2011. Google-indeksin lisääminen malliin parantaa lama-ajalle tehtyjä nykyhetken ennusteita 15 prosenttia.³⁴ Lähitulevaisuuden ennustetarkkuuksia analysoidessaan Tuhkuri havaitsee Google-indeksin sisältävän mallin ennustevirheiden olevan merkittävästi pienempiä kolmen kuukauden päähän tehdyissä ennusteissaan verrattuna nykyhetken vastaaviin. Google-indeksin sisällyttäminen malliin pienentää kolmen kuukauden ennusteen absoluuttista keskivirhettä 39 prosenttia.

³¹ Barreira ym. pyrkivät myös tunnistamaan rakenteelliset muutokset aikasarjoista QLR-testin avulla. Jos aineistossa havaittiin rakenteellinen muutos, aineisto pilkottiin tästä kohdasta ja ainoastaan rakenteellisen muutoksen jälkeisiä havaintoja hyödynnettiin.

³² Barreira ym. havaitsevat kuitenkin Google-indeksien sisältävän mallin kykenevän selittämään osittain automyyntien varianssia. Tämä tulos havaitaan vahvana Portugalille ja heikompana Espanjalle.

³³ Google-indeksin Tuhkuri (2014) rakentaa sanoista: työttömyyskorvaus, työttömyyspäiväraha, ansiopäiväraha, peruspäiväraha, ansiosidonnainen päiväraha ja työttömyyskassa.

³⁴ Tuhkuri havaitsee jopa neljän kuukauden takaisten Google-indeksin viiveiden parantavan nykyhetken ennusteita.

Tuhkurin tulokset indikoivatkin, että Google-indeksin merkitys saattaisi olla jopa suurempi lähitulevaisuuden kuin nykyhetken ennusteille.

Fondeur ja Karamè (2013) tutkivat puolestaan Google-indeksin kykyä parantaa Ranskan työttömyyden ennusteita.³⁵ He rajaavat työttömyyden tarkastelun 15-24-vuotiaiden työttömyyteen, sillä Google-indeksit edustavat heidän mukaansa suurelta osin nuoremman väestön käyttäytymistä. Fondeur ja Karamè käyttävät mallinnuksessaan levitettyä kalman suodinta (*diffuse Kalman filter*) ja suurimman uskottavuuden estimointimenetelmää. Tila-avaruusmalleja (*state-space-models*) käyttämällä he pystyvät hyödyntämään aineistojen eri frekvenssejä nykyhetken ja lähitulevaisuuden ennusteissaan. Heidän käyttämänsä työttömyysaineisto on kuukausitasoista, kun taas Google-hakuaineisto on viikkotasosta.

Fondeur ja Karamè havaitsivat, että Google-indeksin lisääminen tuottaa parhaimmat ennusteet, kun ennuste tehdään viikkoa ennen työttömyystilaston julkistusta. Heidän tuloksiansa mukaan Google-indeksin lisääminen ennustemalliin parantaa ennustetarkkuutta jopa 27 prosenttia ja parannus on poikkeuksetta tilastollisesti merkitsevä. Lisäksi Google-indeksin sisältävien mallien tuottamien ennusteiden keskihajonta laskee keskiarvolla mitattuna 40 prosenttia. Fondeur ja Karamè korostavat myös, että havaitut tulokset olivat heikompia, kun tarkasteltavana oli iäkkäämpien, 25-49-vuotiaiden, työttömyys. Tällöin ennustetarkkuuden parannus vaihteli 9,7-17,5 prosentin välillä. Fondeurin ja Karamén tulokset indikoivatkin Google-hakuaineiston sisältävän jonkinasteista valikoitumisharhaa.

Google-hakuaineistoa on hyödynnetty myös osakemarkkinoilla sijoittajien käyttäytymisen analysointiin. Da ym. (2011) pyrkivät mittaamaan sijoittajien kiinnostusta osakkeeseen rakentamalla Google-indeksejä Russell 3000-osakeindeksin osakkeiden kaupankäyntitunnuksista. He testaavat tutkimuksessaan Barberin ja Odeanin (2008) hypoteesiä, jonka mukaan osakkeeseen osoitetun kiinnostuksen kasvu johtaa positiiviseen hintapaineeseen. He analysoivat Barberin ja Odeanin hypoteesia muuntamalla Google-indeksin muotoon, jossa sen arvo on määritelty vähentämällä tämän viikon arvosta edellisen kahdeksan viikon mediaani. Tämä kuvaa heidän mukaansa epänormaalia hakuintensiteettiä (ASVI-indeksi).

Da ym. tutkivat aluksi tavallisten Google-indeksiensä suhdetta perinteisiin osakekiinnostuksen mittareihin, kuten uutisiin ja yrityksen liikevaihtoon. He havaitsivat Google-indeksien korreloivan ainoastaan hieman perinteisten osakkeen kiinnostusmittareiden kanssa. Lisäksi Granger-kausalisuustestin perusteella Google-indeksien havaitaan olevan edeltäviä muuttujia muihin kiinnostusmittareihin nähden. Da ym. päätulokset tukevat Barberin ja Odeanin hypoteesiä. He havaitsivat kasvun kuluvaan viikon ASVI-indeksissä aiheuttavan osakkeen ylisuoriutumista seuraavan kahden seuraavan viikon aikana. Nämä tulokset ovat huomattavasti vahvempia osakkeille, joilla kauppaa käyvät

³⁵ Heidän käyttämänsä Google-indeksi muodostuu sanasta "Emploi", joka ranskankielessä viittaa sekä sanaan työ että työllisyys.

pääasiassa yksityissijoittajat. Tulokset ovat myös robusteja muiden osakkeen kiinnostusta mittavien mittareiden huomioimiselle.

Edellä esitellyt tutkimukset puoltavat lähes poikkeuksetta Google-hakuaineiston käytön hyödyllisyyttä taloudellisten muuttujien nykyhetken tai lähitulevaisuuden arvoja ennustettaessa. Positiivisia tuloksia on havaittu useassa eri maassa ja usealle eri taloudelliselle muuttujille. Täytyy kuitenkin muistaa, että Google-hakuaineiston hyödyntäminen empiirisessä tutkimuksessa on vielä verrattain nuori ilmiö. Lisäksi ekonometrisen mallintamisen näkökulmasta Googlen tarjoama aineisto kattaa vielä melko lyhyen aikavälin. Tähän astisten tutkimusten päämääränä onkin usein ollut ainoastaan havainnollistaa Google-hakuaineiston mahdollisuuksia yksinkertaisilla tutkimusasetelmilla. Jo tähän mennessä tutkimukset ovat kuitenkin onnistuneet osoittamaan Google-hakuaineiston sisältävän hyödynnettävissä olevaa uudenlaista informaatiota, jonka mahdollisuuksia taloustieteellisessä ennustamisessa ei ole syytä sivuuttaa.

3.5.2 Google-hakuaineisto asuntomarkkinoiden indikaattorina

Google-hakuaineiston mahdollisuuksia on tutkittu myös jonkin verran asuntomarkkinoilla. Tutkimuksissa on havaittu Google-indeksien kykenevän indikoimaan asuntomarkkinoiden suuntaa ja käännteitä. Voimakkaita selityssuhteita on havaittu niin asuntojen hinnoille kuin asuntokauppojen lukumäärällekin. Google-hakuaineistosta muodostetun indikaattorin on myös havaittu parantavan merkittävästi ennustemallin tarkkuutta ja päihittävän asuntomarkkinoiden indikaattorina tutkimuksissa analysoidut perinteiset asuntomarkkinoiden indikaattorit.

Mclaren ja Shanbhogue (2011) tutkivat isobritannialaisella aineistolla, voisiko asuntohintojen nykyarvoja ennustavien AR-mallien ennustetarkkuutta parantaa lisäämällä ennustemalliin Google-indeksin.³⁶ Mclaren ja Shanbhogue testaavat useita eri asuntomarkkinoihin liittyviä hakusanoja päätyen lopulta sanaan "estate agents". Tähän sanaan he päätyvät sanan ja asuntojen hintojen välisen vahvan korrelaation sekä "estate agents"-sanon pienen mittausvirheen vuoksi.

Mclaren ja Shanbhogue havaitsevat suuremman Google-indeksin arvon indikoivan korkeampia asuntojen hintoja, eli toisin sanoen muodostetussa Google-indeksissä asuntokysyntä dominoi asuntotarjontaa. Tutkimuksessaan he myös analysoivat Google-indeksin kykyä selittää asuntojen hintojen kehitystä verrattuna kahteen perinteisempään asuntojen hintojen indikaattoriin, HBF (*Home Builders Federation*) ja RICS (*Royal Institution of Chartered Surveyors*). Google-indeksin lisääminen asuntojen hintoja selittävään AR-malliin parantaa mallin selityssastetta ja pienentää Akaiken informaatiokriteeriä enemmän kuin kahden muun indikaattorin lisääminen. Parhaiten aineistoon sopii kuitenkin malli, jossa AR-mallia on täydennetty sekä Google-indeksillä että HBF- ja RICS-indikaattorilla.

³⁶ Mclaren ja Shanbhogue tutkivat myös Google-indeksien kykyä parantaa työttömyyttä ennustavan AR-mallin ennustetarkkuutta. He havaitsevat, että Google-indeksin huomioiminen mallissa parantaa ennustetarkkuutta enemmän kuin kuluttajaluottamusindikaattori, mutta vähemmän kuin työttömyystukia hakevien henkilöiden lukumäärää kuvaava muuttuja.

Tosin tässä mallissa ainoastaan Google-indeksi on tilastollisesti merkitsevä. Nykyhetken ennustetarkasteluissa McLaren ja Shanbhogue havaitsevat parhaiten asuntojen hintojen nykyarvoja ennustavan mallin sisältävän kaikki tarkasteltavat indikaattorit. Yksittäisistä indikaattoreista Google-indeksin kuitenkin havaitaan parantavan asuntojen hintojen ennustetarkkuutta selvästi eniten. Kokonaisuutena McLarenin ja Shanbhoguen tulokset osoittavat Google-indeksin olevan merkittävämpi asuntojen hintakehitystä kuvaava indikaattori kuin HBF- tai RICS-indikaattori.

Hohenstatt ym. (2011) tarkastelevat Google-indeksien merkitystä kuukausittaisia asuntokauppoja ja -hintoja selittävien mallien aineistoon sopivuudelle. Lisäksi he pyrkivät analysoimaan tuloksiansa suhteessa NARin raportoimaan asunnon ostoprosessiin.³⁷ Tutkimuksessa asuntomarkkinoiden dynaamisia suhteita kuvataan VAR-mallin avulla, jonka endogeenisina muuttujina käsitellään Granger-kausalisuustestin tulosten perusteella asuntomarkkinamuuttujia sekä Googlen "real estate agency"-kategoriaa sekä "apartments"-sanasta muodostettua Google-indeksiä. Lisäksi VAR-malliin lisätään eksogeenisina muuttujina Googlen "home financing"-kategoria sekä makrotaloudellista tilannetta kuvaamaan S&P 500-osakeindeksi, käytettävissä olevat tulot, työttömyysaste sekä asuntolainan korko.

Hohensattin ym. tulosten mukaan Google-indeksien huomioiminen mallissa parantaa asuntojen hintoja selittävän mallin selitysastetta, mutta tulkinta vaikutuksen suunnasta ei ole selkeä, sillä Google-indeksien eri viiveet saavat erimerkkisiä etumerkkejä. Tutkijoiden mielestä tämä johtuu tutkitusta aikavälistä, 1:2004–12:2009, jolloin Yhdysvaltojen taloudellinen tilanne vaihteli voimakkaasti. Muiden heidän tulostensa mukaan erityisesti "real estate agency"-kategoria toimii vakaana indikaattorina asuntokauppojen lukumäärälle. Tosin tulkinta vaikutuksen suunnasta ei ole myöskään nyt täysin selkeä erimerkkisten etumerkkien vuoksi. Lisäksi "real estate agency"-kategorian vaikutusaika asuntokauppojen lukumäärälle noudattelee saadun viivepituuden perusteella NARin raportoimaa ostoprosessikuvausta. He toteavat "real estate agency"-kategorian vaikuttavan asuntokauppojen lukumäärän kautta myös asuntojen hintojen arvoihin. Eksogeenisenä kohdellun "home financing"-kategorian he puolestaan tulkitsevat indikoivan asuntolainauksen kehitystä varsin hyvin. "Home financing"-kategorian kehityksen havaitaan olevan negatiivisessa yhteydessä asuntojen hintoihin ja positiivisessa yhteydessä asuntokauppojen lukumäärään.

Hohenstatt ja Käsbauer (2014) laajentavat Hohenstattin ym. (2011) tutkimusta perehtymällä "real estate agency"- ja "home financing"-kategorioiden merkitykseen Iso-Britannian asuntomarkkinoilla. Lisäksi Hohenstatt ja Käsbauer kiinnittävät tutkimuksessaan enemmän huomiota "home financing"-kategoriaan ja muodostavat kategoriasta asuntomarkkinoiden suhdanteita ennakoivan stressi-indikaattorin, jonka tulisi heidän mukaansa indikoida asuntovelallisten rahoitusvaikeuksien kasvamista. Stressi-indikaattori muodostetaan jakamalla "home financing"-kategorian hakuindeksin arvo myönnettyjen uusien asuntolainojen lukumäärällä.

³⁷ NAR (*National Association of Realtors*) on kansallinen kiinteistövälittäjäyhdistys.

Hohenstatt ja Käsbauer (2014) saavat samankaltaisia tuloksia kuin Hohenstatt ym. (2011). "Real estate agency"-kategorian havaitaan olevan robusti indikaattori myös Iso-Britannian asuntokauppojen lukumäärälle. Tulkinta kategorian vaikutuksen suunnasta ei tosin ole selkeä myöskään isobritannialaisella aineistolla, sillä kategorian etumerkit saavat edelleen sekä positiivisia että negatiivisia etumerkkejä. Hohenstatt ja Käsbauer havaitsevat myös, että "real estate agency"-kategorian kyky indikoida asuntokauppojen lukumäärää heikkenee merkittävästi laskusuhdanteen aikana. Toisen tutkitun kategorian, "home financing", ja siitä muodostetun stressi-indikaattorin he havaitsevat vaikuttavan pääasiassa negatiivisesti sekä asuntojen hintoihin että kauppojen lukumäärään. Stressi-indikaattorin merkityksestä saadut tulokset tutkijat kiteyttävät tulkintaan, jonka mukaan stressi-indikaattorin saadessa arvon 1 se indikoi asuntojen hintojen laskevan seuraavan kuukauden aikana 2,1 prosenttia ja asuntokauppojen lukumäärän laskevan noin 10 prosenttia.³⁸ Stressi-indikaattorin merkitys on siis suurempi asuntokauppojen lukumäärälle kuin asuntojen hinnoille.

Yhdysvaltalaisten asuntomarkkinoiden ennustamisessa Google-hakuaineistoa ovat hyödyntäneet myös Wu ja Brynjolfsson (2013). He keskittyvät neljännesvuositasoisten asuntojen kauppalukumäärän ja hintatason ennustamiseen 50 Yhdysvaltojen osavaltiossa. He käyttävät perusmallinaan asuntomarkkinamuuttujan omia viiveitä sisältävää AR-mallia, jota täydennetään toisen asuntomarkkinamuuttujan ensimmäisellä viiveellä sekä kausivaihtelun ja osavaltiokohtaiset tekijät kontrolloivilla "dummy"-muuttujilla.³⁹ Tutkimuksensa Google-indekseinä he käyttivät Googlen valmiita luokittelukategorioita, joista he valitsivat lopulliseen malliin kaksi alakategoriaa, "real estate listings" ja "real estate agencies".

Wun ja Brynjolfssonin tulosten mukaan erityisesti asuntokauppojen ja Google-indeksien välillä vallitsee voimakas positiivinen korrelaatio. Nykyhetken arvoja ennustaessaan he havaitsevat, että Google-indeksien lisääminen malliin pienentää asuntokauppojen lukumäärän nykyhetken ennusteiden absoluuttista keskivirhettä 2,3 prosenttia ja asuntohintojen absoluuttista keskivirhettä 2,45 prosenttia. Heidän tulostensa mukaan Google-indeksien merkitys on kuitenkin suurempi yhden askeleen päähän ennustettaessa. Asuntomarkkinamuuttujista tämä tulos on vahvempi asuntokauppojen lukumäärälle. Asuntokauppojen lukumäärän yhden neljänneksen ennustetarkkuuden havaitaan paranevan 7,1 prosenttia, kun taas asuntojen hinnoille parannus on huomattavasti pienempi ollen 2,96 prosenttia. Lisäksi Wu ja Brynjolfsson vertaavat Google-hakuaineistolla täydennetyin asuntokauppojen lukumäärän ennustemallin yhden askeleen ennusteita NARin julkaisemaan ennusteeseen ja havaitsevat, että heidän mallinsa ennustetarkkuus on noin 24 prosenttia tarkempi kuin NARin.⁴⁰

Google-hakuaineistoa on sovellettu myös kaupunkien välisen hintakehityksen tutkimiseen. Beracha ja Wintoki (2013) hyödyntävät Google-

³⁸ Stressi-indikaattori saa arvon 1, kun stressi-indikaattorin arvo suurempi kuin 80. prosentti.

³⁹ Molempia asuntomarkkinamuuttujia mallinnetaan omilla malleillaan.

⁴⁰ NAR(National Association of Realtors) on kansallinen kiinteistövälittäjäyhdistys.

aineistoa yhdysvaltalaisen kaupunkien asuntohintakehityksen erojen tutkimiseen. He tutkivat yli 200 yhdysvaltalaisen kaupungin asuntohintojen kehitystä sekä niiden välisiä poikkileikkauseroja. Heidän selitettävänä muuttujanaan on tietyn kaupungin epänormaali hintatuotto (*abnormal price return*), joka on laskettu vähentämällä kaupungin hintojen kasvuasteesta koko markkinan kasvuaste. Beracha ja Wintoki muodostavat jokaiselle kaupungille omat erilliset Google-indeksit. Indeksit kaupungille *i* he muodostavat hakutermeistä "real estate *i*" ja "rent *i*".⁴¹ Kaupunkikohtaisten Google-indeksien kasvuasteista he vähentävät ilman kaupunkirajausta poimitun Google-indeksin kasvuasteen saaden näin epänormaalia hakuintensiteettiä (*abnormal search intensity*) kuvaavan suureen.

Beracha ja Wintoki havaitsivat kaupungin nimellä haetun "real estate *i*"-indeksin ylisuuren hakuintensiteetin arvon aiheuttavan ylisuuren hintatuoton tässä kaupungissa.⁴² Ylisuuri hakuintensiteetin arvo aiheuttaa suurimpia ylisuuria tuottoja kaupungeissa, joissa on joustamaton tonttimaan tarjonta. Tulokset ovat lisäksi robusteja asuntomarkkinasuhtanteen ja väestön kasvun kontrolloinnille.⁴³ Beracha ja Wintoki tutkivat epänormaalien Google-indeksien arvojen sekä epänormaalien asuntojen hintatuottojen suhdetta myös Grangerin kausaalisuusteilla käyttäen VAR-mallissa yhdestä neljään viivettä. Epänormaalit Google-indeksit Granger aiheuttavat jokaisella testatulla viivepituudella epänormaaleja hintatuottoja 1 % merkitsevyystasolla, kun taas epänormaalit hintatuotot Granger aiheuttavat epänormaalien Googlen hakuintensiteetin arvoja tilastollisesti merkitsevästi ainoastaan viiveillä yksi, kolme ja neljä. Granger-kausalisuuden havaitaan siis olevan kahdensuuntaista. Tulos on yhdenmukainen Hohenstättin ym. (2011) ja Hohenstättin ja Käsbauerin (2014) tulosten kanssa.

Dietzel ym. (2014) tutkivat puolestaan, voiko Google-hakuaineiston avulla parantaa liikekiinteistöjen (toimistot, varastot, jne.) kauppohenkilöiden ja hintojen ennusteita.⁴⁴ Liikekiinteistömarkkina koostuu pääasiassa ammattimaisista sijoittajista, mikä tekee sen luonteeltaan poikkeavaksi verrattuna asuntomarkkinaan, jossa ovat mukana myös asumiseen käytetyt kiinteistöt. Google-aineiston hyödyntämisen ongelmaksi saattaa muodostua juuri tämä markkinoiden luonteeseen liittyvä tekijä, sillä on vaikeaa osoittaa, käyttävätkö ammattimaiset sijoittajat tutkimustyössään hakupalvelinta vai onko heillä vakiintuneet verkostot tai tietyt verkkosivut, josta he hakevat suoraan tietoa. Dietzelin ym. mukaan yritysten tutkimusosastot hyödyntävät tiedonhaussa merkittävässä määrin Googlea. He perustelevat näkemystään liikekiinteistöjen ostoprosessia kuvaavilla tutkimuksilla, joissa on osoitettu myös ammattimaisten sijoittajien hyödyntävän internethakuja alkuvaiheen tiedonhaussa (ks. esim. Henderson & Cowart 2002).

⁴¹ Indeksi Miamille on esimerkiksi "real estate Miami" ja "rent Miami".

⁴² Beracha ja Wintoki eivät havaitse merkittävää suhdetta "rent"-sanasta muodostetun kaupungin Google-indeksin ja hintatuoton välillä.

⁴³ Vaihtelut ovat voimakkaampia kuitenkin nousukausien aikana.

⁴⁴ Dietzel ym. (2014) tutkivat liikekiinteistömarkkinan kokonaiskehitystä eivätkä huomioi alueellista vaihtelua millään tavoin.

Dietzel ym. muodostavat seitsemän eri Google-indeksiä. Nämä indeksit koostuvat sekä tutkijoiden valitsemista sanakombinaatioista että Googlen tarjoamista valmiista luokittelukategorioista. Aluksi he tutkivat Google-indeksien suhdetta kappalukumääriin ja asuntojen hintoihin Granger-kausalisuustestien avulla. Granger-kausalisuusteisteistä he havaitsivat Google-indeksien Granger aiheuttavan asuntomarkkinamuuttujia, mutta myös asuntomarkkinamuuttujien Granger aiheuttavan Google-indeksejä. Dietzel ym. korostavat kuitenkin, että Google-indeksien arvot Granger aiheuttavat asuntomarkkinamuuttujien arvoja voimakkaammin kuin toisinpäin. Kausalisuustesteissä havaitusta muuttujien endogeenisestä luonteesta johtuen Google-indeksejä ja asuntomarkkinamuuttujia mallinnetaan VAR-mallilla, jota täydennetään lisäksi yleistä taloudellista tilannetta kuvaavilla eksogeenisillä makrotaloudellisilla muuttujilla.

Dietzelin ym. tulosten perusteella Google-indeksejä sisältävät ennustemallit ovat poikkeuksetta yhden kuukauden ennusteissaan tarkempia olipa vertailtuna perusmallina pelkkiä asuntomarkkinamuuttujia sisältävä malli tai malli, jota on asuntomarkkinamuuttujien lisäksi täydennetty eksogeenisillä makrotaloutta kuvaavilla muuttujilla. Google-indeksien lisääminen parantaa enemmän asuntojen hintoja kuin asuntokauppojen lukumäärää ennustavien mallien ennustetarkkuutta. Tulokset ovat robusteja myös kaikille liikekiinteistösegmenteille (vähittäiskaupat, toimistot ja teollisuushallit sekä koulutustilat). Dietzel ym. toteavat kuitenkin, että tarkimmat ennusteet liikekiinteistömarkkinoista saavutetaan sisällyttämällä ennustemalliin Google-aineisto sekä makrotaloudelliset muuttujat.

Google-hakuaineistoa on hyödynnetty asuntomarkkinoiden analysointiin tutkimalla myös muita muuttujia kuin asuntokauppojen lukumäärää ja asuntojen hintoja. Askitas ja Zimmermann (2014) keskittyvät tutkimuksessaan tarkastelemaan erityisesti asuntomarkkinoiden käännekohtia käyttäen selitettävänä muuttujanaan maksuvaikeuksisten asuntovelallisten osuutta kaikista asuntovelallisista. Google-hakuaineiston avulla he pyrkivät selvittämään, onko asuntovelallisten ongelmia mahdollista havaita reaaliaikaisemmin. Jos tällainen on mahdollista, subprime-kriisin puhkeaminen olisi voinut olla mahdollista ennakoida Google-aineiston avulla. Askitas ja Zimmermann käyttävät Google-indekseinä yksittäisiä sanoja, joita ovat "hardship letter", "short sale" ja "real estate owned".

Askitaksen ja Zimmermannin tulosten mukaan heidän rakentamansa Google-indeksit onnistuvat tunnistamaan asuntovelallisten maksukyvyssä tapahtuneet muutokset. Kun he lisäävät "hardship letter"-sanasta muodostetun Google-indeksin perusmallina käyttämäänsä kausitasoitettuun AR-malliin, ennustemallin tarkkuus paranee huomattavasti. Lisäksi he havaitsivat "short sale"- ja "real estate owned"-sanoista muodostettujen indeksien kykenevän ennustamaan muutaman viikon tarkkuudella finanssikriisin seurauksena Yhdysvaltain hallituksen käynnistämän pankkien pelastusohjelman, TARPin (*Troubled Assets Rescue Program*).

TAULUKKO 3 Google-hakuaineisto ja asuntomarkkinat

Tekijä	Tutkittava muuttuja	Google-hakuaineiston lisäämisen vaikutus
Mclaren & Shanbhogue (2011)	Asuntohintojen kehitys Isossa-Britanniassa.	"Estate agents"-sanasta muodostetun Google-indeksin sisältämä malli on ennustetarkkuudeltaan selvästi tarkempi verrattuna malleihin, jossa on käytetty kahta muuta asuntomarkkinoiden indikaattoria.
Hohenstatt ym. (2011)	Yhdysvaltalaisen asuntojen hinnat ja kauppoyöjen lukumäärät.	Googlen "real estate agency"-kategoria on robusti indikaattori asuntokaupoille. "Home financing" -kategoria kykenee puolestaan indikoimaan jossain määrin markkinoiden asuntolainatilannetta.
Hohenstatt & Käsbauer (2014)	Asuntojen hinnat ja kauppoyöjen lukumäärät Iso-Britanniassa.	Googlen "real estate agency"-kategoria on robusti indikaattori asuntokaupoille. "Home financing" -kategoriasta rakennettu stressi-indikaattori kykenee kuvaamaan asuntomarkkinamuuttujien, erityisesti kauppoyöjen lukumäärän, kehitystä.
Wu & Brynjolfsson (2013)	Asuntokauppoyöjen lukumäärän ja asuntojen hintoyöjen kehitys 50 eri yhdysvaltalaisessa osavaltiossa.	Google-indeksit parantavat enemmän asuntokauppoyöjen ennusteita. Vaikutus on voimakkain yhden neljänneksen päähän ennustettaessa. Lisäksi Google-indeksin sisältävä kauppoyöjen lukumäärän ennustemalli on tarkempi ennusteissaan kuin NARin ennuste.
Beracha & Wintoki (2013)	Asuntojen hintoyöjen erot yhdysvaltalaisen kaupunkien välillä.	"Real estate i"-indeksin ylisuuri hakuintensiteetin arvo aiheuttaa ylisuuren hintatuoton tässä kaupungissa.
Dietzel ym. (2014)	Yhdysvaltalaisen liikekiinteistöjen hinnat sekä kauppoyöjen lukumäärät.	Google-aineiston sisältämät mallit ovat yhden kuukauden ennusteissaan selvästi tarkempia kuin perusmallit. Kun Google-indeksti lisätään ennustemalliin, paranee ennustetarkkuus enemmän liikekiinteistöjen hinnoille kuin kauppoyöjen lukumäärälle.
Askatas & Zimmermann (2014)	Maksuvaikeuksisten asuntovelallisten määrä suhteessa kaikkiin asuntovelallisiin.	Google-aineiston kyetään havainnoimaan paremmin muutokset maksuvaikeuksisten talouksien lukumäärässä. Lisäksi Google-aineistolla olisi kyetty ennustamaan tarkasti TARP:n alkuaikojä.

4 EMPIIRINEN MENETELMÄ

Tässä luvussa käydään läpi empirian toteutukseen liittyvä teoriatausta. Google-hakuaineiston ja asuntomarkkinamuuttujien välisten suhteiden mallintamisen lähtökohdaksi valittu empiirinen menetelmä pohjautuu esimerkiksi Hohenstattin ym. (2011) ja Dietzelin ym. (2014) asuntomarkkinoiden kuvaamiseen soveltamaan vektoriautoregressiiviseen (VAR) malliin, jossa asuntojen hintoja, asuntokauppojen lukumääriä ja Google-indeksejä käsitellään simultaanisina endogeenisina muuttujina. Tähän malliin lisätään yleistä taloudellista tilaa kontrolloivia eksogeenisiä makrotaloudellisia muuttujia.

Aiemmissä tutkimuksissa (ks. esim. Hohenstatt 2011 ym.) asuntomarkkinamuuttujia ja Google-hakuindeksejä on mallinnettu VAR-mallilla, jotta muuttujien mahdolliset simultaaniset suhteet kyetään havaitsemaan. Tutkijat ovat olettaneet, että vaikka Google-hakuindeksien havaittasiinkin selittävän asuntomarkkinoiden muuttujia, on täysin mahdollista, että myös asuntomarkkinamuuttujat selittävät Google-hakuindeksien kehitystä. Tämä simultaaninen suhde onkin saanut tukea myös kirjallisuuden tuloksista (Hohenstatt ym. 2011, Hohenstatt & Käsbauer 2014 ja Dietzel ym. 2014).⁴⁵ Asuntomarkkinoita mallintavassa kirjallisuudessa on myös usein käsitelty asuntomarkkinamuuttujien ja makrotaloudellisten muuttujien välistä suhdetta endogeenisena (ks. esim. Oikarinen 2012). Tässä tutkimuksessa makrotaloudelliset muuttujat käsitellään kuitenkin eksogeenisina, sillä tutkimuksen ensisijaisena tavoitteena on selvittää, kyetäänkö asuntomarkkinoiden kehitystä indikoimaan paremmin lisäämällä Google-indeksit malliin.

Empiirisen menetelmän esittelyssä käydään ensiksi läpi yleisimpien yksikköjuuritestien teoria. Sen jälkeen keskustellaan, miksi mahdolliset rakenteelliset muutokset pitäisi kontrolloida aineistosta, ja esitellään Zivotin ja Andrews (1992) yksikköjuuritestin menetelmä rakenteellisten muutosten tunnistamiseksi. Tämän jälkeen käydään läpi VAR-mallin perusteoria, jonka jälkeen kerrotaan, miksi muuttujien mahdollinen yhteisintegroituvuus tulee tutkia ennen lopullisen mallin spesifiointia. Lopuksi tarkastellaan tutkimuksessa sovellettuja ennustetarkkuuden mittareita. Teoriaosuus pohjautuu suurelta osin Brooks (2008) teokseen.

4.1 Yksikköjuuritestit

Yksikköjuuritesteillä tutkitaan muuttujien stationaarisuutta, joka on edellytys robusteille tulkinnoille muuttujien välisistä suhteista. Aikasarjan sanotaan olevan

⁴⁵ Asuntomarkkinamuuttujien ja Google-indeksien simultaanista suhdetta on tutkittu kirjallisuudessa Granger-kausalisuustesteillä.

stationaarinen, jos sillä on ajan suhteen vakio keskiarvo (kaava 2), äärellinen varianssi (kaava 3) sekä autokovarianssirakenne (kaava 4).

$$E(y_t) = \mu, \quad (2)$$

$$E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty, \quad (3)$$

$$E(y_t - \mu)(y_{t_2} - \mu) = \gamma_{t_2-t_1} \quad \forall t_1, t_2 \quad (4)$$

Usein taloudellisten muuttujien aikasarjassa on kuitenkin niin sanottu yksikköjuuri. Tällä tarkoitetaan sitä, että aikasarjan havainnot generoivan prosessin karakteristisen yhtälön juuri on yksi. Yksikköjuuri aiheuttaa prosessin epästationaarisuuden, eli prosessiin kohdistuvilla sokeilla on pysyvä vaikutus aikasarjan kulkuun, eikä sillä silloin ole esimerkiksi ajassa vakiota keskiarvoa. Esimerkki yksikköjuuren sisältävästä prosessista on satunnaisen kulun prosessi (*random walk*).

Muuttujien epästationaarisuus vaikeuttaa muuttujien välisten suhteiden tulkintaa. Pahimmillaan epästationaaristen muuttujien käyttö mallinnuksessa johtaa täysin vääriin tulkintoihin muuttujien välisistä suhteista. Epästationaaristen muuttujien käyttö analyysissä saattaa johtaa tilastollisesti merkitsevien regressiokerrotoimien ja ylisuuren selityksasteen, R^2 , havaitsemiseen, vaikka todellisuudessa muuttujat olisivatkin toisistaan täysin riippumattomia. Kahden muuttujan välisissä tarkasteluissa tämä on mahdollista esimerkiksi silloin, kun molemmat muuttujat kasvavat tietyllä kasvuvauhdilla eli niiden trendikehitys on samanlainen.

Stationaarisuuden testaamiseen on kehitetty useita erilaisia yksikköjuuritestejä. Eräs yleisesti käytetty yksikköjuuritesti on Dickey-Fuller-yksikköjuuritesti (Dickie & Fuller 1979).⁴⁶ Yksikköjuuritesti pohjautuu kaavan 5 yhtälöön.

$$\Delta y_t = \rho y_{t-1} + u_t \quad (5)$$

jossa Δ on differenssioperaattori eli $\Delta y_t = y_t - y_{t-1}$. Kaavan 5 yhtälön parametrin ρ estimointi suoritetaan pienimmän neliösumman menetelmällä. Dickie-Fuller-testin nollahypoteesina on, että prosessilla on yksikköjuuri, eli toisin sanoen $\rho = 0$. Yksisuuntainen vastahypoteesi on, että $\rho < 0$. Muuttujan todetaan olevan stationaarinen, kun nollahypoteesi hylätään. Regressoitavaan DF-testin yhtälöön on mahdollista lisätä myös parametreja kontrolloimaan aikasarjan deterministisiä komponentteja, trendiä tai "driftiä".⁴⁷ Tällöin kaavan 5 estimoitava yhtälö muuntuu kaavan 6 mukaiseen muotoon.

$$\Delta y_t = \rho y_{t-1} + \mu + \gamma t + u_t \quad (6)$$

jossa μ on vakiokomponentti ja t on deterministinen trendikomponentti, jonka kerroin γ on. Kun stationaarisuutta testataan kaavan 6 yhtälöllä, selvitetään, onko aikasarja stationaarinen sekä vakio- että trendikomponentin suhteen.

Empiirisissä tutkimuksissa DF-testin sijasta käytetään yleensä sen laajennettua versiota, ADF-testiä (*Augmented Dickey-Fuller-test*). ADF-testi ottaa huomioon

⁴⁶ Dickie-Fuller-testiin viitataan tästä eteenpäin DF-testinä.

⁴⁷ "Driftillä" tarkoitetaan komponenttia, joka pysyy ajan suhteen vakiona.

myös virhetermin u_t mahdollisen autokorreloituneisuuden. Estimoitava lineaarinen yhtälö on esitetty kaavassa 7.

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + u_t \quad (7)$$

Kaavassa 7 muuttujaa Δy_t selitetään y_t :n kertaalleen viivästetyn arvon lisäksi myös Δy_t :n viiveillä. Kaavaan 7 voi lisätä vakio- ja trendikomponentin samoin kuin esitettiin DF-testille kaavassa 6. ADF-testin optimaalisen viiverakenne voidaan määrittää useilla tavoilla. Yksinkertaisena lähtökohtana viivepituuden valintaan voidaan pitää aineiston frekvenssiä. Jos käytettävä aineisto on esimerkiksi kuukausittaista, tulisi yhtälöön sisällyttää 12 viivettä. Usein empiriassa ADF-testin optimaalinen viivepituus määritetään kuitenkin informaatiokriteereiden avulla (ks. luku 4.3).

Sekä Dickie-Fuller- ja laajennetun Dickie-Fuller-testin nollahypoteesina on, että aikasarjan epästationaarisuus, minkä vuoksi estimaateista lasketut tilastolliset testisuureet eivät noudata perinteistä Studentin t -jakaumaa. Parametriestimaattien tilastollista merkitsevyyttä testatakseen Fuller (1976) simuloi Monte Carlo-menetelmällä DF- ja ADF-testin kriittiset arvot. Kun testisuureen suuruus on enemmän negatiivinen kuin DF- ja ADF-testin kriittiset arvot, nollahypoteesi ei saa tukea ja aikasarjan todetaan olevan stationaarinen.

Edellä esitellyt yksikköjuuritestit eivät ole täysin ongelmattomia. Eräs keskeinen DF-testejä kohtaan esitetty kritiikki on niiden heikko tilastollinen voimakkuus silloin, kun tarkasteltava aikasarja sisältää lähes yksikköjuuren, eli selittävän viiveen kerroinparametri on lähellä arvoa yksi, mutta kuitenkin pienempi.⁴⁸ Tällöin aikasarja on todellisuudessa stationaarinen, mutta systeemiin kohdistuva sokki kuolee hyvin hitaasti. Empiirisissä tarkasteluissa DF-testit ovat osoittautuneet voimakkuudeltaan heikoiksi etenkin pienten otosten tapauksissa, joissa tarkasteltava aikasarja on lähes yksikköjuuri-prosessi.

4.2 Rakenteellisten muutosten tunnistaminen

Monissa taloudellisissa aikasarjoissa on havaittu aikasarjojen havainnot generoivaa prosessia muuttaneita rakenteellisiä muutoksia. Rakenteellisesta muutoksesta puhutaan, kun aikasarjassa tapahtuu sokin vaikutuksesta diskreetti muutos tietyllä ajanhetkellä. Sokki voi aiheuttaa aikasarjaan myös hitaan pitkän aikavälin kuluessa etenevän muutoksen. Rakenteellinen muutos voi tapahtua aikasarjan vakio- tai trendikomponentissa tai molemmissa yhtäaikaan, eli toisin sanoen sokin seurauksena aikasarjan keskiarvo, varianssi tai autokorrelaatorakenne ei säily ajassa vakiona. Esimerkki tietynä ajanhetkenä tapahtuvasta rakenteellisestä muutoksesta on Bretton Woodsin valuuttajärjestelmän romahtaminen vuonna 1972, mikä aiheutti rakenteellisen

⁴⁸ Tilastollisella voimakkuudella tarkoitetaan sitä todennäköisyyttä, jolla testi hylkää nollahypoteesin silloin, kun nollahypoteesi ei ole totta.

muutoksen Yhdysvaltain dollarin ja Englannin punnan väliseen valuuttakurssiainkasarjaan.

Rakenteellisten muutosten läsnäolo vaikeuttaa muuttujien analysointia. Rakenteelliset muutokset aikasarjassa vaikeuttavat esimerkiksi perinteisten DF-yksikköjuuritestien tulkintaa. Perron (1989) korostaa, että jos rakenteellista muutosta ei huomioida yksikköjuuritestauksessa, kyky hylätä väärä yksikköjuuritestin nollahypoteesi heikkenee. Näin ollen muuttuja todetaan epästationaariseksi, vaikka se todellisuudessa onkin stationaarinen. Kirjallisuudessa onkin kehitelty useita yksikköjuuritestejä, jotka kykenevät testaamaan aikasarjan stationaarisuutta rakenteellisen muutoksen läsnäollessa (ks. esim. Perron 1989 ja Zivot & Andrews 1992).

Rakenteelliset muutokset aikasarjassa ovat este myös lineaaristen mallien hyödyntämiselle muuttujien välisten suhteiden kuvaamisessa.⁴⁹ Esimerkiksi rakenteellisen muutoksen käsittävälle aikavälille estimoitu lineaarinen regressiomalli tuottaa vääriä tulkintoja, sillä se olettaa virhetermien keskiarvon ja varianssin pysyvän samana koko tarkastelujaksolla, eli toisin sanoen se ei kykene huomioimaan muutoksia vakiokomponenteissa tai muuttujien kasvuvauhdeissa. Lineaarisen mallin estimoinnista saadut muuttujien regressiokertoimet kertovat ainoastaan muuttujien välisestä keskimääräisestä suhteesta tarkasteluperiodin ajalta. Lineaarisia malleja on kuitenkin mahdollista soveltaa rakenteellisen muutoksen sisältävälle aikavälille, jos aineistosta muodostetaan kaksi erillistä osaotosta pilkkomalla aineisto rakenteellisen muutoksen kohdalta.⁵⁰ Ongelmaksi muodostuu kuitenkin usein otosten pienempi koko ja siitä johtuva tehokkuustappio.

Kuten edellä todettiin, rakenteellisten muutosten tunnistaminen on tärkeää robustien tulkintojen aikaansaamiseksi. Joissain tutkimuksissa rakenteellisten muutosten ajankohta on valittu tutkijan ennakkotietämyksen perusteella (ks. esim. Perron 1989). Tällaista eksogeenistä rakenteellisten muutoksen tunnistustapaa on kirjallisuudessa pidetty usein aineiston louhintana, jonka tavoitteena on saada aikaiseksi tietynlaisia empiirisiä tuloksia (ks. esim. Zivot ja Andrews 1992). Rakenteellisten muutosten tunnistamiseen onkin kehitetty useita endogeenisiä menetelmiä, eli toisin sanoen rakenteellisen muutoksen ajankohta määritellään algoritmin avulla. Tässä tutkimuksessa mahdollinen rakenteellinen muutos pyritään tunnistamaan Zivotin ja Andrews (1992) kehittämällä endogeenisellä menetelmällä.

Zivot ja Andrews (1992) kehittivät yksikköjuuritestin, jolla voidaan testata aikasarjan stationaarisuutta sallimalla yksi rakenteellinen muutos joko muuttujan vakio- tai trendikomponentissa tai vaihtoehtoisesti molemmissa. Rakenteellisen muutoksen ajankohta määritellään endogeenisesti peräkkäisten testien avulla hyödyntämällä koko tutkimusaineistoa ja testaamalla jokaiselle mahdolliselle rakenteellisen muutoksen ajanhetkelle omaa "dummy"-muuttujaa. Rakenteellisen muu-

⁴⁹ DF-testit ovat myös lineaarisia testejä.

⁵⁰ Toinen mahdollinen tapa on rakenteellisen muutoksen kohtelemisen "dummy"-muuttujana.

toksen ajankohdaksi valitaan se havainto, jonka ollessa "dummy"-muuttujana vakio- tai trendikomponentin tai molemmat sisältävän ADF-testin t-testisuure minimoituu, eli on eniten negatiivinen. Lopulliseksi rakenteellisen muutoksen ajankohdaksi valitaan siis se ajanhetki, jolloin yksikköjuuritestin nollahypoteesi, eli se, että aikasarjassa on yksikköjuuri, saa vähiten tukea.

4.3 VAR-mallin teoria

VAR-mallin ekonometrisenä popularisoijana pidetään talouden nobelisti Christopher Simsiä (1980). Hän generalisoi yhden muuttujan aikasarjamallista monta muuttujaa sisältävän VAR-mallin, joka pystyy käsittelemään useita selitettäviä muuttujia ja huomioimaan näiden mahdollisen simultaanisen luonteen eli sen, että muuttujan nykyhetken arvoon saattavat vaikuttaa toisen selitettävän muuttujan arvot ja toisinpäin. VAR-malleja on sovellettu erityisesti taloudellisten ja rahoituksen ilmiöiden dynaamisten suhteiden tutkimiseen.

Yksinkertaisimmillaan VAR-malli voidaan esittää kahden muuttujan systeeminä, jossa muuttujien y_{1t} ja y_{2t} tämän ajanhetken arvot riippuvat molempien muuttujien viiveiden kombinaatioista ajanhetken k asti sekä virhetermeistä u . Kaavat 8 ja 9 havainnollistavat kahden muuttujan VAR-systeemiä.

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \dots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \dots + \alpha_{1k}y_{2t-k} + u_{1t} \quad (8)$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \dots + \beta_{2k}y_{2t-k} + \alpha_{21}y_{1t-1} + \dots + \alpha_{2k}y_{1t-k} + u_{2t} \quad (9)$$

jossa u_{it} on valkoisen kohinan virhetermi ominaisuuksilla: $E(u_{it}) = 0$, $i = 1, 2$ ja $E(u_{1t}u_{2t}) = 0$.

Edellä esitetty kahden muuttujan VAR-malli voidaan esittää usealla eri tavalla. Tarkastellaan esitystapoja yksinkertaisessa tapauksessa, jossa k eli viivepituus saa arvon yksi. Yhden viiveen VAR-malli on esitetty yhtälömuotoisena kaavoissa 10-11 ja matriisimuotoisena kaavassa 12.⁵¹

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + u_{1t} \quad (10)$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + u_{2t} \quad (11)$$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (12)$$

VAR-mallia käytettäessä muuttujien tulee olla stationaarisia. Jos muuttujien havaitaan olevan epästationaarisia, ne täytyy muuntaa stationaarisiksi ennen estimointien aloittamista. Yleisin tapa muuntaa muuttujat stationaariksi on

⁵¹ Tutkimuksen empiriaan sovellettuja VAR-malleja kuvataan kaavan 12 matriisimuotoisella esityksellä.

käyttää niiden ensimmäisiä differenssejä.⁵² Joskus myös muuttujien lineaarinen kombinaatio saattaa olla stationaarinen, vaikka muuttujat itsessään olisivatkin epästationaarisia. Muuttujien yhteinen stokastinen trendi eli muuttujien yhteisintegroituvuus täytyykin tutkia aina VAR-mallia rakennettaessa. Jos muuttujien havaitaan olevan keskenään yhteisintegroituneita, tulee virheenkorjaustermi lisätä VAR-malliin ja estimoinnit suorittaa käyttäen vektorimuotoista virheenkorjausmallia(VECM). Muuttujien välisen yhteisintegraation tunnistaminen käydään tarkemmin läpi seuraavassa kappaleessa.

Rajoittamatonta VAR-mallia käytettäessä (edellä kaavoissa 8-12 esitetty tilanne) jokaisella yhtälön endogeenisella muuttujalla tulee olla sama viivepituus. Viiveiden pituudet VAR-mallissa on mahdollista määrittää käyttämällä esimerkiksi monimuuttujayhtälöille laajennettuja informaatiokriteerejä. Kaavoissa 13-15 on esitetty monimuuttujayhtälöille laajennetut Akaiken, Schwarz-Bayesin ja Hannan-Quinnin informaatiokriteerit.

$$MAIC = \log|\hat{\Sigma}| + 2k'/T \quad (13)$$

$$MSBIC = \log|\hat{\Sigma}| + \frac{k'}{T}\log(T) \quad (14)$$

$$MHQIC = \log|\hat{\Sigma}| + \frac{2k'}{T}\log(\log(T)) \quad (15)$$

jossa $\hat{\Sigma}$ on jäännöstermien varianssi-kovarianssi-matriisi, T on havaintojen lukumäärä ja k' regressoreiden kokonaislukumäärä kaikissa yhtälöissä yhteensä. k' määräytyy yhtälöstä $p^2k + p$, jossa p on yhtälöiden lukumäärä ja k viiveiden lukumäärä kussakin yhtälössä lisättynä vakiolla. VAR-mallin optimaalisen viivepituuden arvo on se, jolle informaatiokriteerin arvo minimoituu. Informaatiokriteerit voivat kuitenkin antaa eri ehdotuksen optimaalisesta viiverakenteesta. Tällöin tutkijan tulee valita, mitä kriteeriä hän ensisijaisesti suosii. Tässä tutkimuksessa VAR-mallin optimaalisen viivepituuden määrittämisessä käytetään ensisijaisesti Schwarz-Bayesin(MSBIC) informaatiokriteeriä.⁵³

Rajoittamattomassa VAR-mallissa kaikkia muuttujia käsitellään endogeenisina. Systeemiin on kuitenkin mahdollista lisätä myös eksogeenisia muuttujia. Tällöin puhutaan VARX-mallista, joka itseasiassa on eräänlainen rajoitettu VAR-malli. VARX-mallissa oletetaan valittujen eksogeenisten muuttujien yhtälöiden oikean puolen muuttujien kertoimet nolliksi, eli toisin sanoen eksogeenisten muuttujien arvot määräytyvät VAR-systeemin ulkopuolella.

VAR-mallissa muuttujien välisiä suhteita on yleensä vaikea tulkita keskittymällä ainoastaan muuttujille estimoituihin kertoimiin. Lisäksi näitä

⁵² Muuttujat voidaan muuntaa stationaariseksi myös esimerkiksi estimoimalla aikasarjan trendi- ja kausivaihtelukomponentit ja poistamalla sitten nämä komponentit havainnot generoivasta prosessista.

⁵³ Schwarz-Bayesin informaatiokriteeri on sama kuin Bayesilainen informaatiokriteeri ja se lyhennetään tässä tutkimuksessa aina sanalla BIC.

kertoimia saattaa usein olla todella monta.⁵⁴ Eräs tapa analysoida muuttujien keskinäisiä suhteita on tarkastella niitä Granger-kausalisuustestin avulla. Kausalisuustestin kehittäjä Sir Clive Granger vuoden 1969 artikkelissaan, minkä jälkeen siitä on tullut yleisesti käytetty testi muuttujien välisiä dynaamisia suhteita analysoitaessa. Granger-kausalisuustestillä tutkitaan VAR-systeemissä, onko muuttujilla toisiaan tai itseään selittävää kykyä. VAR-mallin muuttujien kertoimien selittävää kykyä testataan F -testisuurella, jonka nollahypoteesina on, että muuttujalla ei ole selittävää kykyä selitettävänä olevaan muuttujaan.

Kaavoissa 16-17 on esitetty kahden viiveen VAR-malli kahden muuttujan systeemille. Granger kausalisuutta analysoitaessa molempien muuttujien yhtälöissä F -testisuure lasketaan erikseen kummallekin muuttujalle, y_1 ja y_2 , huomioiden muuttujien molempien viiveiden kertoimet. F -testisuure testaa siis yhteishypoteesia muuttujien, y_1 ja y_2 , kertoimille.

$$y_{1t} = \alpha_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + \gamma_{11}y_{1t-2} + \gamma_{12}y_{2t-2} + u_{1t} \quad (16)$$

$$y_{2t} = \alpha_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + \gamma_{21}y_{1t-2} + \gamma_{22}y_{2t-2} + u_{2t} \quad (17)$$

Jos F -testisuureen yhteisnollahypoteesi hylätään muuttujan y_1 kertoimille muuttujan y_2 yhtälössä (kaava 17), muuttujan y_1 todetaan Granger aiheuttavan muuttujaa y_2 . Muuttujan y_1 viiveet ovat siis kokonaisuutena merkitseviä muuttujan y_2 yhtälössä. Tämä ei välttämättä tarkoita sitä, että kaikki y_1 muuttujan kertoimet olisivat merkitsevästi nolasta eroavia. Jos muuttujan y_2 havaitaan lisäksi Granger aiheuttavan muuttujaa y_1 , puhutaan yleisesti kahdensuuntaisesta kausaaliiteetista. Lopuksi täytyy todeta, että Granger-kausalisuustesti ei välttämättä kerro mitään todellisesta kausalisuussuhteesta. Se kertoo ainoastaan, edeltääkö toinen muuttuja toista tai itseään tai ovatko muuttujien liikkeet yhtäaikaista.

4.4 Yhteisintegraation tutkiminen Johansenin menetelmällä

Kuten VAR-malleista keskustellessa todettiin, muuttujien yhteisintegraatio täytyy tutkia ennen lopullisen mallin spesifiointia. On mahdollista, että VAR-malliin valituilla muuttujilla on sama stokastinen trendi, jonka huomioiminen mallinnuksessa on välttämätöntä oikeiden tulkintojen aikaansaamiseksi. Samalla stokastisella trendillä tarkoitetaan sitä, että pitkällä aikavälillä muuttujien stokastiset trendit käyttäytyvät identtisesti, eli toisin sanoen muuttujien keskinäinen suhde (etäisyys toisiinsa) säilyy ajassa vakiona. Jos muuttujien havainnot generoivien prosessien havaitaan sisältävän saman stokastisen trendin, puhutaan yhteisintegraatiosta.

⁵⁴ Jos VAR-systeemissä on g määrä muuttujia, joista jokaisella k määrä viiveitä, täytyisi mallissa analysoida $g+kg^2$ parametria. Kolmen muuttujan ja kolmen viiveen mallilla analysoitavia parametreja kertyisi esimerkiksi 30.

Yhteisintegroituvuus edellyttää vähintään kahden muuttujan olevan epästationaarisia samalla integraation asteella. Muuttujan integraatioasteella tarkoitetaan sitä differenssien lukumäärää, joka muuttujasta on otettava stationaarisuuden saavuttamiseksi. Esimerkiksi jos muuttujan y_t havaitaan olevan tasomuuttujana epästationaarinen, mutta muuttujan ensimmäisen differenssin havaitaan olevan stationaarinen, puhutaan asteella yksi integroituneesta prosessista. Asteella yksi integroitunut prosessi voidaan esittää myös muodossa $I(1)$. Vaikka kaksi muuttujaa olisivatkin erikseen tarkasteltuina $I(1)$ -prosesseja, voi niiden lineaarinen kombinaatio olla stationaarinen eli $I(0)$ -prosessi, kun muuttujien välillä on sama stokastinen trendi. Tästä muuttujien välisestä stationaarisesta lineaarikombinaatiosta käytetään yleisesti nimeä virheenkorjaustermi. Muuttujien ollessa yhteisintegroituja virheenkorjaustermi tulee sisällyttää estimoitavaan malliin.

Eräs tapa tutkia muuttujien välistä yhteisintegraatiota on Johansenin testi (Johansen 1988 ja Johansen & Juselius 1990). Johansenin testillä voidaan testata samanaikaisesti useampaa kuin yhtä yhteisintegraatiosuhdetta. Näin ollen sitä voidaan hyödyntää yhteisintegraatiotarkasteluissa, joissa on mukana enemmän kuin kaksi muuttujaa.⁵⁵ Johansenin testausmenetelmässä perinteisen pienimmän neliösumman menetelmän sijasta estimointi tapahtuu suurimman uskottavuuden estimointimenetelmällä.

Johansenin yhteisintegraatiotesti pohjautuu VAR-mallista johdettuun pitkän aikavälin vektorimuotoiseen virheenkorjausmalliin (VECM). VECM-malli on esitetty kaavassa 18 g määrälle kertaalleen differoituja muuttujia (yhtälön vasen puoli), joita selitetään VECM-systeemiin valittujen differenssimuotoisten muuttujien viiveillä (yhtälön oikea puoli), joista jokaiselle viivepituudelle kohdistetaan oma kerroinmatriisi Γ . Π voidaan tulkita muuttujien pitkän aikavälin kerroinmatriisiksi.

$$\Delta y_t = \Pi y_{t-k} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t \quad (18)$$

jossa $\Pi = (\sum_{i=1}^k \beta_i) - I_g$, $\Gamma_i = (\sum_{j=1}^i \beta_j) - I_g$ ja u_t ovat Gaussilaisia virhetermejä keskiarvolla nolla sekä varianssimatriisilla Σ .

Johansenin yhteisintegraatiotestauksessa keskitytään tarkastelemaan nimenomaan pitkän aikavälin kerroinmatriisia Π . Pitkällä aikavälillä Γ kerroinmatriisit katoavat, sillä pitkän aikavälin tasapainossa muuttujien Δy_{t-i} arvot ovat nolliä. Lisäksi virhetermit u_t voidaan asettaa odotusarvoonsa eli arvoon nolla. Näin ollen kaavan 18 yhtälöstä jää pitkällä aikavälillä tarkasteltavaksi ainoastaan $\Pi y_{t-k} = 0$.

Muuttujien välisen yhteisintegraation tutkiminen perustuu matriisiin Π asteeseen, jota tarkastellaan matriisin ominaisarvojen kautta.⁵⁶ Muuttujien välisten yhteisintegraatiovektoreiden lukumäärää voidaan analysoida kahdella matriisin

⁵⁵ Perinteisellä Engle-Granger-menetelmällä (Engle & Granger 1987) on mahdollista tutkia samanaikaisesti ainoastaan kahden muuttujan välistä yhteisintegraatiota.

⁵⁶ Matriisin Π aste on yhtä suuri kuin nollassa eroavien karakteristen juurien (ominaisarvojen) lukumäärä. Matriisin Π aste on myös yhtä suuri kuin yhteisintegraatiovektoreiden lukumäärä r .

Π ominaisarvoihin perustuvalla testisuurella: λ_{trace} ja λ_{max} . Näiden testisuureiden laskenta on esitetty kaavoissa 19 ja 20.

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^g \ln(1 - \hat{\lambda}_i) \quad (19)$$

$$\lambda_{max}(r, r + 1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (20)$$

jossa r kuvaa nollahypoteesissa oletettujen yhteisintegraatiovektoreiden lukumäärää ja $\hat{\lambda}_i$ on estimoitu ominaisarvo matriisista Π . λ_{trace} testisuureen nollahypoteesi on, että yhteisintegraatiovektoreiden määrä on yhtä suuri tai pienempi kuin r . Vastahypoteesina λ_{trace} testissä on, että muuttujien välillä on enemmän kuin r yhteisintegraatiovektoria. λ_{max} testisuureen nollahypoteesina puolestaan on, että yhteisintegraatiovektoreiden lukumäärä on yhtä suuri kuin r . Vastahypoteesina λ_{max} testissä on, että yhteisintegraatiovektoreiden lukumäärä on $r+1$. Kun mainittujen testisuureiden nollahypoteesia ei enää hylätä eli estimoitu testisuureen arvo ei ole testin kriittistä arvoa suurempi, päädytään lopulliseen muuttujien välillä olevien yhteisintegraatiovektoreiden lukumäärään.

Jotta λ_{trace} ja λ_{max} testisuureiden hypoteeseja kyetään analysoimaan, niille täytyy estimoida kriittiset arvot. Testisuureiden jakauma ei noudata perinteistä Studentin t -jakaumaa, minkä vuoksi Johansen ja Juselius (1990) määrittävät testisuureiden kriittiset arvot. Nämä kriittiset arvot riippuvat estimoitavien muuttujien kokonaismäärästä g ja testattavien yhteisintegraatiovektoreiden, r , lukumäärästä, eli toisin sanoen kriittiset arvot riippuvat epästationaaristen komponenttien lukumäärästä. Lisäksi testisuureiden kriittisiin arvoihin vaikuttaa se, onko mallien estimoinneissa huomioitu vakio- ja trendikomponentteja.

4.5 Ennustetarkkuuden tarkastelu

Ennustetarkkuuden mittaamiseen valittiin kaksi eri tarkkuusmittaria: keskineliövirhe(MSE) sekä Theilin (1966, 1971) U1 -kerroin.⁵⁷ Näiden mittareiden avulla kyetään tekemään tulkintoja Google-hakuaineiston merkityksestä ennustemallin tarkkuudelle.

Keskineliövirheen(MSE) laskenta perustuu ennustevirheeseen, jonka laskenta on esitetty kaavassa 21.

$$\hat{e}_{t+1,t} = A_{t+1} - F_{t+1} \quad (21)$$

jossa A_{t+1} on todellinen toteutunut arvo aikasarjalle ajanhetkellä $t+1$ ja F_{t+1} yhden askeleen ennuste aikasarjalle, joka on estimoitu ajanhetkellä t . MSE on ennustevirheiden neliöiden keskiarvo (kaava 22).

⁵⁷ Dietzel ym. (2014) käyttävät samoja ennustetarkkuuden mittareita asuntomarkkinoiden ennustamista Googlen hakuaineiston avulla käsittelevässä artikkelissaan.

$$MSE = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 \quad (22)$$

Toisen valitun ennustetarkkuuden mittarin, Theilin U1-kertoimen, laskenta perustuu osittain keskineliövirheeseen. Theilin U1-kerroin vaihtelee yleensä arvojen 0 ja 1 välillä ja sen laskenta on esitetty kaavassa 23.⁵⁸

$$U1 = \frac{\sqrt{MSE}}{\sqrt{\frac{1}{n} \sum_{i=1}^n A_i^2 + \frac{1}{n} \sum_{i=1}^n F_i^2}} \quad (23)$$

jossa A kuvaa jälleen todellisia toteutuneita aikasarjan arvoja ja F kuvaa ennustemallilla saatuja ennusteita.

Molempien mittareiden tuottamien arvojen tulkinta on samankaltainen. Mitä pienemmän arvon mittari saa, sitä parempi on testatun mallin ennustetarkkuus. Mittareiden arvoja suoraan vertailemalla voidaan todeta, kumpi ennustemalleista tuottaa tarkempia ennusteita.

⁵⁸ Jos Theilin U1-kerroin on yli 1, tulkitaan mallin ennuste arvaamista huonommaksi.

5 AINEISTO

Tutkimuksen aineistona käytetään Tilastokeskuksen StatFin-tietokannan avointa tilastoaineistoa sekä *Google Trends* palvelun tarjoamaa hakuaineistoa. Aineisto sisältää 43 neljännesvuosittaista havaintoa käsittäen aikavälin vuoden 2004 ensimmäisestä neljänneksestä vuoden 2014 kolmanteen neljännekseen. Analysoitavista muuttujista käytetään sekä nimellisiä että reaalisia muuttujia. Pää tarkastelussa keskitytään nimellisiin muuttujiin, mutta tulosten robustisuustarkastelussa asuntojen hinnat, käytettävissä olevat tulot henkeä kohti sekä asuntolainan korko muunnetaan reaalisiksi Tilastokeskuksen kuluttajahintaindeksin avulla.

5.1 Asuntomarkkinoiden aineisto

Asuntomarkkinoita kuvaamaan valittiin Tilastokeskuksen säännöllisesti julkaisema tilasto koko Suomen maantieteellisen alueen käsittävistä vanhojen osakehuoneistojen hinnoista sekä niiden kauppalukumääristä.⁵⁹ Vanhoista osakehuoneistoista rivitalot rajattiin kokonaan tarkastelun ulkopuolelle, minkä seurauksena tarkasteluun jäivät ainoastaan vanhojen kerrostalojen osakehuoneistot käsittäen kerrostalo-osakkeet aina yksiöistä kolmioita suurempiin. Rivitalot rajattiin tarkastelusta, sillä tarkasteltavaksi ryhmäksi haluttiin valita yksi mahdollisimman homogeeninen ryhmä, jolla käydään runsaasti ja säännöllisesti kauppaa (ks. Oikarinen 2012). Lisäksi kerrostaloasunnot muodostavat jo itsessään merkittävän osan, 44 prosenttia, Suomen asuntokannasta (Tilastokeskus 2014b).⁶⁰

Vanhojen osakeasuntojen hinnat -tilasto kuvaa osakeasuntojen velattomien myyntihintojen kehitystä. Hinnoista käytetään koko maan käsittävää kausitasoitamatonta nimellistä hintaindeksisarjaa, jonka perusvuotena on vuosi 2000. Hintatiedon lisäksi Tilastokeskus raportoi tietoja myös Suomessa toteutuneiden vanhojen vapaarahoitteisten kerrostalo-osakkeiden kauppalukumääristä. Vapaarahoitteiseksi asunnoksi Tilastokeskus määrittelee asunnon, joka ei ole ARA- eikä HITAS-asunto. Edellytyksenä asuntokaupan näkymiselle Tilastokeskuksen tilastossa ovat myös Verohallinnolle tehty varainsiirtoilmoitus toteutuneesta asuntokaupasta sekä tiedot rakennuksen valmistumisvuodesta, asunnon koosta ja postinumerosta. (Tilastokeskus 2015b.)

Hinta- ja asuntokauppatilasto saatiin suurelta osin kerättyä Tilastokeskuksen avoimesta StatFin-tietokannasta. Tietokanta ei kuitenkaan sisältänyt tietoja asuntokauppojen lukumäärästä vuosilta 2004–2007. Tiedot näiden vuosien asuntokaupoista minulle toimitti Petri Kettunen Tilastokeskukselta.

⁵⁹ Tilastokeskus määrittelee vanhan osakehuoneiston asunnoksi, joka ei ole valmistunut tarkasteluvuonna eikä sitä edeltävänä vuonna (Tilastokeskus 2015b).

⁶⁰ Suomen asuntokannasta rivitalojen osuus on noin 13,5 prosenttia (Tilastokeskus 2014b).

5.2 Google-aineisto

Vanhojen kerrostaloasuntojen markkinoita indikoimaan rakennetaan kaksi eri Google-indeksiä. Indeksit muodostetaan yksittäisistä sanoista omaan subjektiiviseen ennakkotietämykseen perustuen.⁶¹ Google-indeksejä poimittaessa rajausaikavälinä oli 1:2004–1:2015.⁶² Indeksien maantieteellisenä rajauksena käytettiin puolestaan koko Suomea. Vaikka Suomessa asuntomarkkinat eroavatkin alueellisesti, valitulla tarkasteluajavälillä Google-hakuaineistoa ei ollut mahdollista rajata maantieteellisesti tarkemmin. Vasta vuodesta 2013 lähtien Google-indeksejä on ollut saatavilla koko Suomea tarkemmilla maantieteellisillä rajauksilla. Tutkimuksessa tiedostetaan, että tarkemman maantieteellisen rajauksen puuttuminen saattaa aiheuttaa Google-indekseihin myös jonkinasteista alueellista valikoitumista. Esimerkiksi indekseihin valittujen hakusanojen käyttötavassa saattaa olla alueellisia eroja, minkä seurauksena Google-indeksien arvot edustaisivat enemmän tiettyjen alueiden ihmisten hakukäyttäytymistä.⁶³ Hakuaineiston kuvailun yhteydessä korostettiin myös Google-indeksien sisältävän muutaman prosentin mitausvirheen. Tästä aiheutuva kohinaa pyrittiin kontrolloimaan ottamalla kahden otospäivän aritmeettinen keskiarvo Google-indeksien arvoista.

Ensimmäinen indeksi (Google1-indeksi) rakennettiin kuvaamaan asuntomarkkinoihin kohdistuvaa kiinnostusta hyvin yleisellä tasolla.⁶⁴ Indeksistä muodostettiin hakutermitä: ”asunto+asunnot+myytävä asunto+myytävät asunnot”.⁶⁵ Indeksistä kontrolloitiin lisäksi vuokra-asuntoja koskevat haut rajaamalla indeksin ulkopuolelle vuokra-asumiseen viittaavat sanat.⁶⁶ Jos vuokra-asumiseen liittyviä sanoja ei olisi kontrolloitu, indeksin ja asuntomarkkinoiden välinen suhde olisi lähtökohtaisesti ollut vieläkin vaikeampi tulkita, koska indeksin hakuintensiteetin kasvu olisi voinut viestiä myös vuokra-asumiseen kohdistuvan kysynnän kasvusta. On kuitenkin mahdollista, että vuokra-asuntoa etsivä henkilö on voinut kirjoittaa myös pelkän asunnot-sanalla ilman minkäänlaista viitettä vuokra-asumisesta, minkä seurauksena indeksi saattaa edelleen sisältää informaatiota vuokra-

⁶¹ Yksittäisistä sanoista rakennetuilla indekseillä asuntomarkkinoita ovat kuvanneet myös esimerkiksi McLaren ja Shanbhogue (2011).

⁶² Google-aineistona käytettiin havaintoja syyskuuhun 2014 saakka (kolmas vuosineljännes). Aineistoa kerätessä Google-hakuaineisto oli saatavilla tammikuuhun 2015 saakka, minkä vuoksi sitä käytettiin aineiston rajauksessa loppuajankohtana. Rajauksella olisi merkitystä Google-indeksien arvoihin, jos Google-indeksi olisi saanut maksimiarvonsa jonain syyskuun 2014 jälkeisenä viikkona. Näin ei kuitenkaan ollut, joten rajauksella ei ole merkitystä Google-indeksien arvoihin.

⁶³ Jonkinasteisesta alueellisesta valikoitumisesta viestii esimerkiksi sanoista ”etuovi+oikotie+jokakoti” (Google2-indeksi) muodostetun Google-indeksin alueellinen jakautuminen. Hakusanoilla haetaan suhteellisesti eniten Helsingissä, kun taas esimerkiksi Jyväskylässä sanat eivät ole yhtä suosittuja. Tämä voi toisaalta viestiä myös Helsingin vilkkaammista asuntomarkkinoista Jyväskylään verrattuna.

⁶⁴ Google1- ja Google2-indeksin hakutermit ovat kokonaisuudessaan esitettyinä liitteessä 1.

⁶⁵ Hakutermit ”myytävä asunto” sisältyy jo itseasiassa asunto-sanalla tehtyihin hakuihin, joten sitä ei olisi ollut välttämätöntä rajata erikseen. Indeksien valittujen sanojen kanssa samaan aikaan haettiin eniten sanoja: oikotie, oikotie asunnot, etuovi, etuovi asunnot ja asunnot helsinki.

⁶⁶ Näitä sanoja olivat vuokra, vuokra-asunto, vuokrattavat, vuokra asunto ja vuokra-asunnot.

asumiseen liittyvästä kysynnästä. Lisäksi on korostettava, että indeksi sisältää todennäköisesti sanojen yleisestä luonteesta aiheutuvaa kohinaa. Esimerkiksi indeksiin valitulla asunnot-sanalla tietoa hakevan henkilön intressit voivat olla aivan muut kuin asunnon ostoon tai myyntiin liittyvät. Indeksien otospäivinä olivat 17.5.2015 ja 20.5.2015.⁶⁷

Toinen muodostettu Google-indeksi (Google2-indeksi) pyrkii kuvaamaan asuntoportaaleihin kohdistuvaa kiinnostusta. Hakuportaaleiden ovat ensisijainen informaatiolähde asunnon myymistä tai ostamista harkitsevalle henkilölle (Huoneistokeskus 2011). Tämän vuoksi onkin syytä olettaa, että asuntoportaaleihin kohdistuvilla Google-hauilla voidaan indikoida asuntomarkkinamuuttujien kehitystä. Google2-indeksi muodostettiin yhdistämällä kolmen suosituimman suomalaisen asuntoportaalien nimet yhteen indeksiin. Indeksini muodostettiin hakutermitä: "etuovi+oikotie+jokakoti".⁶⁸ Valitut asuntoportaaleiden nimet ovat brändinimiä. Kirjallisuudessa esimerkiksi McLaren ja Shanbhogue (2011) ovat korostaneet, että brändinimien käyttö Google-indeksissä on ongelmallista, sillä niiden suosio muuttuu yleensä ajassa. Suosion vaihtelu pyrittiin kontrolloimaan yhdistämällä Google2-indeksiin Suomen kolme suosituinta asuntoportaalaa ja sallimalla näiden asuntoportaaleiden suosion vaihtelun näiden kolmen portaalien välillä (TNS Gallup 2014). Otospäivinä tälle indeksille olivat 2.3.2015 ja 3.3.2015.⁶⁹

Kuten Google1-indeksiinkin, myös Google2-indeksiin sisältyy kohinaa. Valituista hakuportaaleista on mahdollista etsiä informaatiota muustakin kuin asunnoista. Oikotie.fi-sivustolta voi asuntojen lisäksi etsiä myös työpaikkoja tai autoja. Lisäksi jokaisesta asuntoportaalista on mahdollista etsiä myytävien asuntojen ohella myös vuokra-asuntoja. Myytävät asunnot eivät myöskään rajoitu ainoastaan tämän tutkimuksen kiinnostuksen kohteena oleviin kerrostaloasuntoihin, vaan niiden lisäksi tarjonta käsittää myös muun muassa rivi- ja omakotitalot.⁷⁰ Google2-indeksin käytön ongelmaksi saattaakin muodostua se, että kykeneekö indeksi havainnoimaan riittävällä tarkkuudella kerrostalohuoneiston ostosta kiinnostuneen henkilön käyttäytymistä.

Valittujen Google-indeksien lisäksi testattiin muita mahdollisia asuntomarkkinoihin liittyviä hakusanoja. Tällaisia hakusanoja olivat esimerkiksi kiinteistövälittäjä, kerrostalokaksio, asuntokupla, asuntomarkkinat ja asuntolaina. Näistä sanoista muodostettuja indeksejä ei kuitenkaan valittu, sillä näihin sanoihin kohdistui pääsääntöisesti hyvin vähäinen määrä hakuja, mikä olisi johtanut todennäköiseen suureen mittausvirheeseen. Lisäksi lähes poikkeuksetta nämä hakusanat saivat joinain ajanhetkinä arvokseen nollan (ks. esim. kuvio 3).

⁶⁷ Viikkotasosten Google1-indeksisarjojen keskihajontojen prosenttiero oli 2,7 prosenttia.

⁶⁸ Kaksi ensimmäistä asuntoportaalaa olivat olemassa koko tarkasteluperiodin ajan. Jokakoti.fi palvelu puolestaan perustettiin vuonna 2011. Tästä aiheutui jonkinasteinen hyppäys viikkotasosten indeksissä, mutta sen merkitys pieni olemattomiin aggregoinnin ja kausitasoituksen jälkeen. Jokakoti.fi palvelu haluttiin sisällyttää malliin kontrolloimaan mahdollista substituutiovaikutusta.

⁶⁹ Viikkotasosten Google2-indeksisarjojen keskihajontojen prosenttiero oli 1 prosentti.

⁷⁰ Esimerkiksi Etuovi.com-sivustolla oli tarkasteluaikavälillä kohdemäärällä tarkasteltuna karkeasti yhtä paljon myynti-ilmoituksia kerrostaloista kuin omakotitaloistakin. Rivitalojen myynti-ilmoituksia oli noin puolet kerrostalojen määrästä. (Etuovi.com, markkinapuntari 2015).

Molemmilla rakennetuilla Google-indekseillä oletetaan olevan positiivinen suhde asuntokauppoihin. Kun henkilö hakee indeksiin valittuja sanoja, oletetaan sen olevan indikaatio kasvavasta asuntokysynnästä, joka edelleen johtaa suurempaan kerrostaloasuntojen kauppalukumäärään. Hintojen ja Google-indeksien suhteesta ei tehdä samanlaista lähtöoletusta, sillä hakuintensiteetin kasvu voi liittyä sekä kysyntä- että tarjontapuoleen. Näistä vaikutuksista dominoivampi määrää lopulta Google-indeksien ja hintojen suhteen. McLaren ja Shanbhogue (2011) korostavat, että kysyntä- ja tarjontapuolen erottaminen indeksiä rakennettaessa on erittäin haastellista. He toteavat, että usein on helpompi antaa mallin tulosten paljastaa se, kumpi vaikutuksista on hallitsevampi.

5.3 Makrotaloudellinen aineisto

Jotta asuntomarkkinamuuttujien ja valittujen Google-indeksien suhteesta kyettäisiin tekemään luotettavia tulkintoja, huomioitiin mallia rakennettaessa myös yleistä taloudellista tilannetta kuvaavia makrotaloudellisia muuttujia.⁷¹ Makrotaloudelliset muuttujat oletettiin malleissa eksogeenisiksi. Makrotaloudellisista muuttujista mukaan valittiin työpäiväkorjattu käytettävissä oleva tulo henkeä kohti, asuntolainan korko sekä työttömyysaste. Työttömyysaste valittiin, koska sitä pidetään yleisesti hyvänä indikaattorina vallitsevasta taloudellisesta tilasta ja sitä on käytetty aiemmissakin tutkimuksissa Google-indeksien ja asuntomarkkinamuuttujien välistä suhdetta kuvaavassa VAR-systeemissä (ks. esim. Dietzel ym. 2014). Käytettävissä olevien tulojen henkeä kohti (ks. esim. Case & Schiller 2004) sekä korkojen (ks. esim. Oikarinen 2012) on puolestaan havaittu selittävän asuntomarkkinoiden nykytilaa merkittävässä määrin. Makrotaloudellisista muuttujista kerättiin kausitasoittamattomat tilastot.

⁷¹ Makrotaloudellisten muuttujien aikasarjat poimittiin Tilastokeskuksen StatFin-tietokannasta.

6 TULOKSET

Ennen varsinaisen analysoinnin aloittamista osa muuttujista täytyi muuntaa neljännesvuositasoisiksi, sillä valitulla tarkasteluperiodilla kiinnostuksen kohteena olevat kerrostaloasuntojen hinnat ja kauppojen lukumäärät ovat saatavilla ainoastaan tällä frekvenssillä.⁷² Muuttujien aggregointi päätettiin toteuttaa perinteisellä aritmeettisella keskiarvolla. Ensimmäisiksi aggregoitaviksi muuttujiksi valittiin viikkotasoiset Google-indeksien arvot. Kirjallisuudessa esimerkiksi Hohenstatt ym. (2011) ja Dietzel ym. (2014) ovat korostaneet Google-indeksien sisältävän merkittävää kausivaihtelua. Kausivaihtelun tutkimiseksi Google-indeksien arvot muunnettiin ensiksi kuukausitasolle, jolla muuttujien kausivaihtelua analysoitiin graafisen analyysin perusteella.⁷³ Graafisessa analyysissä kuukausitasoisten Google-indeksien havaittiin sisältävän selvää kausivaihtelua, minkä vuoksi kuukausitasoiset Google-indeksit päätettiin kausitasoittaa. Kausitasoitus suoritettiin muun muassa Hohenstattin ym. (2011) käyttämällä X11 Census-menetelmällä.⁷⁴ Kuukausitasoiset Google-indeksit aggregoitiin kausitasoituksen jälkeen neljännesvuositasolle.

Muista tutkimuksessa käytetyistä muuttujista kaikki muut paitsi asuntolainan korko olivat jo valmiiksi neljännesvuositasoisia. Kuukausitasoinen asuntolainan korko aggregoitiin neljännesvuositasolle aritmeettisellä keskiarvolla. Koska Google-indeksit päätettiin kausitasoittaa, robustien tulkintojen aikaansaamiseksi myös muut tarkastelussa olevat muuttujat oli järkevää kausitasoittaa. Muutkin muuttujat kausitasoitetiin samalla X11 Census-menetelmällä. Suurimpien poikkeuksellisten havaintojen tasoittamiseksi kaikille muille muuttujille paitsi asuntolainan korolle tehtiin myös log-transformaatio.⁷⁵

Valittu tarkasteluaikajakso sisältää myös merkittäviä taloudellisia muutoksia, jotka saattavat aiheuttaa tarkasteltavien muuttujien havainnot generoivaan prosessiin rakenteellisia muutoksia. Tarkasteluaikavälille osuu esimerkiksi vuonna 2008 alkanut globaali finanssikriisi, jonka seurauksena vanhojen asuntojen hinnat putosivat Suomessa 5 prosenttia vuosien 2008 ja 2009 ensimmäisten neljännesten välillä (Suomen Pankki 2011).⁷⁶ Rakenteellisen muutoksen läsnäoloa ja mahdollista ajankohtaa tutkittiinkin tämän vuoksi asuntomarkkinamuuttujille sekä Google-indekseille Zivot-Andrews-testillä (ks. liite 3). Zivot-Andrews-testi ehdotti kaikille muuttujille rakenteellisen muutoksen ajankohdaksi eri ajanhetkeä, kun muuttujien viivepituus määritettiin samalla informaatiokriteerillä. Asuntomarkki-

⁷² Puhuttaessa asuntojen hinnoista ja asuntokaupoista viitataan aina kerrostaloasuntojen hintoihin ja kauppalukumääriin, vaikka sitä erikseen mainittaisikaan.

⁷³ Kuukausitasolle aggregoidessa viikko katsottiin kuuluvaksi siihen kuukauteen, jossa sillä lukumäärällisesti oli enemmän päiviä.

⁷⁴ X11 Census-menetelmällä kausitasoitus tehtiin multiplikatiivisella menetelmällä.

⁷⁵ Kausitasoitettut ja log-transformoidut asuntomarkkinamuuttujat sekä Google-indeksit ovat esitettyinä liitteissä 2 ja 3.

⁷⁶ Mahdollinen rakenteellinen muutos oletettiin tutkimuksessa yksittäiseksi nopeasti ohimeneväksi shokiksi, jolla ei ole pysyvää vaikutusta muuttujan havainnot generoivaan prosessiin.

namuuttujien ja Google-indeksin (Google1- tai Google2-indeksi) VAR-malliin sisällytettiin kuitenkin rakenteellista muutosta kuvaava "dummy"-muuttuja vuoden 2008 toiselle neljännekselle. Tämä ajanhetki oli Zivot-Andrews-testin ehdottama rakenteellisen muutoksen ajankohta sekä asuntojen hinnoille että asuntokauppojen lukumäärälle, tosin eri viivepituuden määrittämiseksi käyttäen. Kyseisen ajanhetken "dummy"-muuttuja ei kuitenkaan ollut yhdessäkään testattujen VAR-mallien yhtälöistä tilastollisesti merkitsevä, joten se päätettiin jättää kontrolloimatta lopullisissa malleissa.⁷⁷

6.1 Aineiston kuvailu

Tasomuotoisten muuttujien stationaarisuutta analysoitiin laajennetulla Dickie-Fuller-testillä (ADF-testi). ADF-testien tulokset sekä taso- että differenssimuuttujille ovat esitettyinä taulukossa 4. Viivepituus ADF-testiin määritettiin Bayesilaisella informaatiokriteerillä. Bayesilaisen informaatiokriteerin ehdottama viivepituus on ilmoitettu suluissa ADF-testin testisuureen jälkeen.

TAULUKKO 4 ADF-testien tulokset

Muuttuja	Tasomuuttujat	Differoidut muuttujat
Asuntokaupat (log) (AK)	-1,87 (0)	-6,71*** (0)
Asuntojen hinnat (log) (AH)	-2,33 (1) (t)	-3,69*** (0)
Google1 (log) (G1)	-1,31 (0)	-6,82*** (0)
Google2 (log) (G2)	-2,80 (0) (t)	-2,67* (1)
Tulo per capita (log) (Y)	-2,35 (0)	-5,58*** (0)
Työttömyysaste (log) (U)	-2,65* (2)	-4,89** (0)
Asuntolainan korko (I)	-1,53 (1)	-3,78*** (0)

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Kaikki tasomuuttujat testattiin lähtökohtaisesti vakiokomponentin sisältävällä ADF-testillä. Jos graafisen tarkastelun perusteella kuitenkin havaittiin muuttujan havainnot generoivan prosessin sisältävän deterministisen trendikomponentin, sisällytettiin se vakion sijasta ADF-testin yhtälöön. Differoidut muuttujat testattiin vakiokomponentin kanssa.

Suomalaisella aineistolla Oikarinen (2012) havaitsee kerrostaloasuntojen hintojen olevan epästationaarinen tasomuuttujana ja kerrostaloasuntojen kauppakumäärän olevan 5 % merkitsevyydellä stationaarinen. Taulukon 4 tulosten perusteella tulkinta kerrostaloasuntojen hintojen epästationaarisuudesta noudattelee Oikarisen (2012) tuloksia. Asuntokauppojen lukumäärälle tulkinta puolestaan eroaa Oikarisen (2012) tuloksista, sillä muuttujan havaitaan olevan tarkastelujaksolla epästationaarinen. Tämä johtunee todennäköisesti tutkimuksen lyhemmästä

⁷⁷ Mahdollista rakenteellista muutosta kontrolloimaan rakennettiin myös lyhyen aikavälin käsittäviä "dummy"-muuttujia. "Dummy"-muuttuja rakennettiin esimerkiksi aikavälille 2:2008-1:2009. Muuttuja oli tilastollisesti merkitsevä kuitenkin ainoastaan Google-indeksiä selittävissä VAR-mallin yhtälöissä. Tämän vuoksi sitä ei käytetty lopullisissa malleissa.

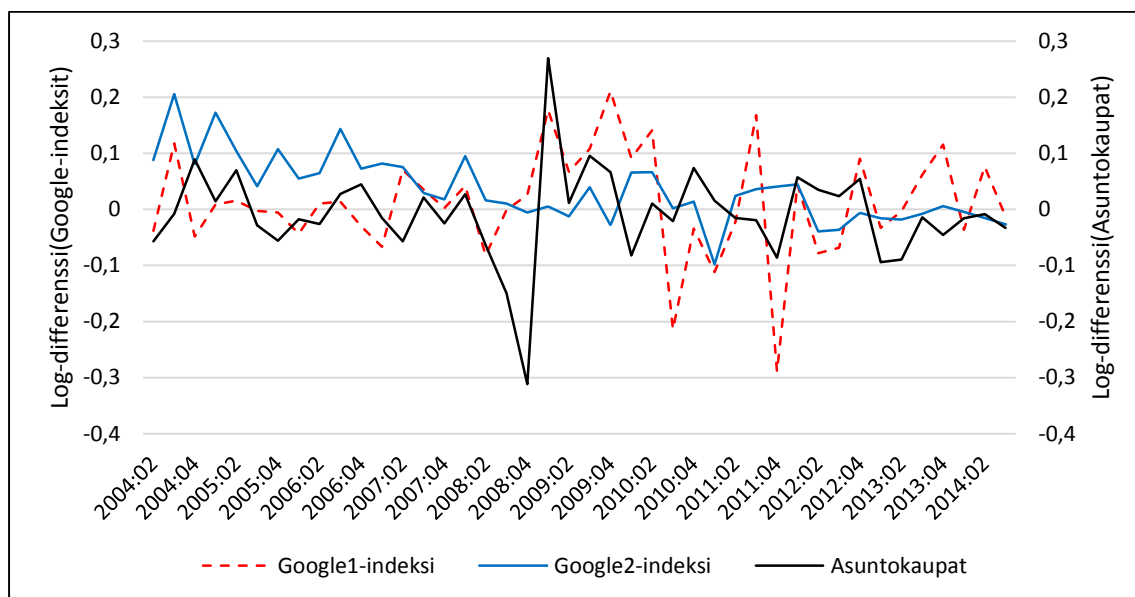
ja turbulenteja taloudellisia aikoja sisältävästä tarkasteluperiodista.⁷⁸ Taulukosta 4 havaitaan myös Google-indeksien olevan epästationaarisia tasomuuttujina, mikä on yhdenmukainen tulos aiemman kirjallisuuden kanssa (ks. esim. Suhoy 2009 ja Hohenstatt ym. 2011). Google2-indeksin tosin havaitaan olevan 1 % merkitsevyystasolla stationaarinen, kun ADF-testin estimoitavaan yhtälöön sisällytetään trendikomponentin sijasta vakiokomponentti. Kokonaisuutena tasomuotoisista muuttujista ainoastaan työttömyysaste on tilastollisesti merkitsevä 10 % merkitsevyystasolla. Stationaarisuuden aikaansaamiseksi kaikki muuttajat differoitiin eli muuttajien nykyhetken arvoista vähennettiin niiden edellisen neljänneksen arvo. Muunnoksen jälkeen kaikki muuttajat olivat stationaarisia vähintään 10 % merkitsevyystasolla.

Kuvioissa 4 ja 5 ovat esitettyinä differoitujen asuntomarkkinamuuttujien ja Google-indeksien aikasarjat. Kuvioista voidaan havaita, että molemmissa asuntomarkkinamuuttujissa on suuri piikki 2008 vuoden viimeisellä neljänneksellä ja vuoden 2009 ensimmäisellä neljänneksellä, kun taas kummallekaan Google-indeksille vastaavanlaista piikkiä ei havaita. Asuntomarkkinamuuttujissa havaittu piikki kuvaa mitä todennäköisimmin globaalien finanssikriisien aiheuttamia vaikutuksia Suomen asuntomarkkinoilla. Google-indeksien muutoksia tarkastelemalla havaitaan Google1-indeksin olevan selvästi volatiilimpi näistä kahdesta.⁷⁹ Erityisen suurta Google1-indeksin vaihtelu näyttäisi olevan alkuvuodesta 2010 loppuvuoteen 2011. Mitään yksittäistä syytä indeksin suurelle vaihtelulle ei voida kuitenkaan luotettavasti esittää. Erot Google-indeksien vaihtelun suuruudessa selittyvät suurimmaksi osaksi indeksien sisältämien sanojen luonteella. Google2-indeksin asuntoportaaleita kuvaavien sanojen suosio on kasvanut tasaisesti koko tarkasteluperiodin ajan asuntomarkkinoiden siirtyessä yhä enemmän verkkoon ja palveluiden keskittyessä yhä enemmän asuntoportaaleihin. Tosin kuvioista 4 ja 5 havaitaan, että Google2-indeksin kasvu on hidastunut koko tarkasteluperiodin ajan ja vuodesta 2013 alkaen kasvu on ollut lähes nollassa. Tämän voidaan tulkita viittaavan siihen, että asuntoportaaleiden nimillä tehdyt haut ovat saavuttaneet jo maksimiarvonsa. Toisaalta se voi myös viitata siihen, että yhä useampi henkilö menee asuntoportaaliin hakupalvelimen käyttämisen sijasta suoraan URL-osoitteella. Toisin kuin Google2-indeksille, Google1-indeksin sanoista ei puolestaan ole mahdollista antaa yhtä suoraviivaista tulkintaa.

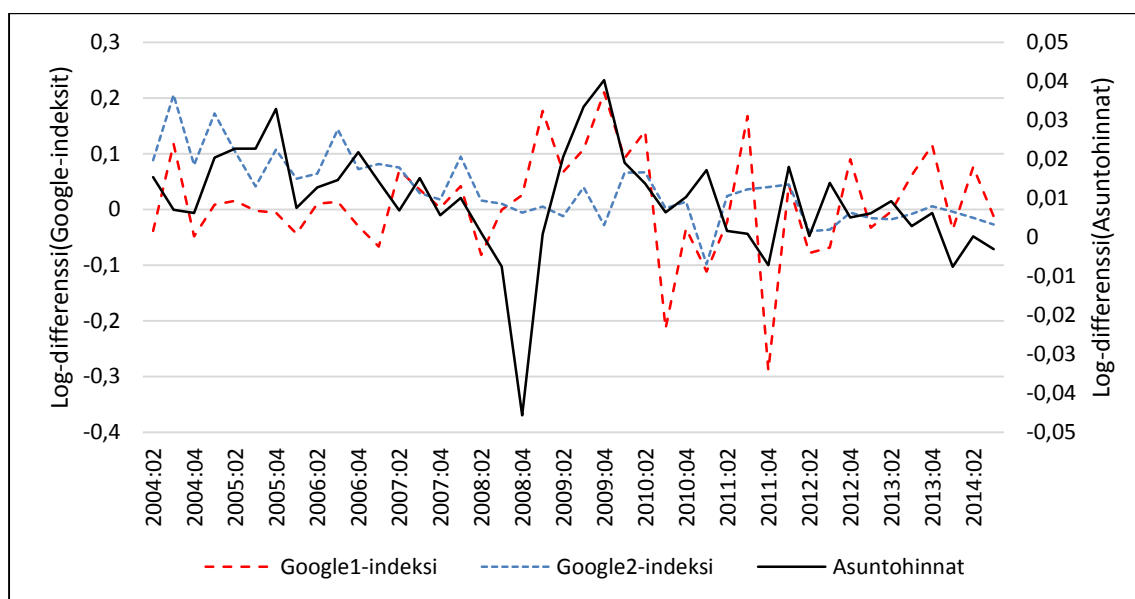
Kuvioista 4 ja 5 voidaan lisäksi karkeasti arvioida asuntomarkkinamuuttujien ja Google-indeksien muutosten välistä edeltäjä-seuraaja-suhdetta (*lead-lag relationship*). Google-indeksien muutosten havaitaan ainakin jossain määrin edeltävän asuntomarkkinamuuttujien muutoksia ennen finanssikriisin ensimmäisen aallon saapumista Suomeen eli ennen vuoden 2008 viimeistä neljännestä. Tämän ajanhetken jälkeen muuttajien välinen suhde vaikuttaisi heikkenevän merkittävästi ja katoavan jopa kokonaan.

⁷⁸ Oikarisen (2012) tarkasteluajavälinä ovat vuodet 1988-2008.

⁷⁹ Google1-indeksin varianssi on yli kaksi kertaa suurempi kuin Google2-indeksin (ks. liite 5).



KUVIO 4 Differoidut Google-indeksit ja asuntokaupat



KUVIO 5 Differoidut Google-indeksit ja asuntojen hinnat

Taulukossa 5 on esitetty differoitujen muuttujien kontemporaaaliset korrelaatiokertoimet. Korrelaatioista havaitaan Oikarisenkin (2012) havaitsema asuntojen hintojen ja kauppalukumäärän muutosten erittäin suuri ja tilastollisesti merkitsevä positiivinen korrelaatio. Asuntomarkkinamuuttujien välillä havaitusta korrelaatiokertoimesta voidaan tulkita, että asuntomarkkinamuuttujien muutokset ovat vahvassa positiivisessa lineaarisessa suhteessa ja että muutokset asuntojen hinnoissa ja kauppalukumäärässä liikkuvat yleensä samaan suuntaan. Taulukosta 5 havaitaan myös, että Google-indeksien muutoksista ainoastaan Google1-indeksin muutokset korreloivat tilastollisesti merkitsevästi asuntokauppalukumäärän

muutosten kanssa. Kummankaan Google-indeksin ei havaita korreloivan tilastollisesti merkitsevästi asuntojen hintojen kanssa. Lisäksi havaitaan, että Google-indeksit eivät korreloi tilastollisesti merkitsevästi keskenään, mikä viestii siitä, että muodostetut Google-indeksit eivät sisällä samanaikaisesti samanlaista informaatiota.

Taulukon 5 tulosten mukaan muutokset asuntomarkkinamuuttujissa ja Google-indekseissä korreloivat myös voimakkaasti makrotalouden muutosten kanssa. Asuntokauppojen lukumäärän muutosten havaitaan korreloivan negatiivisesti ja 1 % tasolla tilastollisesti merkitsevästi asuntolainan koron muutosten kanssa, kun taas asuntohintojen muutokset korreloivat positiivisesti ja 10 % tasolla tilastollisesti merkitsevästi tuloissa (ΔY) tapahtuneiden muutosten kanssa. Asuntojen hintojen kehitys vaikuttaisi siis liikkuvan jossain määrin samansuuntaisesti käytettävissä olevien tulojen muutosten kanssa, kun taas asuntokauppojen lukumäärän muutosten voidaan tulkita kulkevan jossain määrin eri suuntaan korkomuutosten kanssa. Google-indekseistä Google1-indeksin muutosten havaitaan puolestaan korreloivan negatiivisesti ja tilastollisesti merkitsevästi asuntolainan koron muutosten kanssa, kun taas Google2-indeksin muutosten havaitaan korreloivan negatiivisesti ja tilastollisesti merkitsevästi työttömyysasteen muutosten kanssa. Lisäksi havaitaan, että makrotaloudellisten muuttujien muutokset ovat vahvassa lineaarisessa suhteessa toisiinsa korreloiden keskenään vahvasti ja tilastollisesti merkitsevästi.

TAULUKKO 5 Differoitujen muuttujien kontemporaaaliset korrelaatiokertoimet

Muuttuja	$\log(\Delta AK)$	$\log(\Delta AH)$	$\log(\Delta G1)$	$\log(\Delta G2)$	$\log(\Delta Y)$	$\log(\Delta U)$	ΔI
$\log(\Delta AK)$	1						
$\log(\Delta AH)$	0,52***	1					
$\log(\Delta G1)$	0,27*	0,21	1				
$\log(\Delta G2)$	0,07	0,25	0,13	1			
$\log(\Delta Y)$	-0,07	0,31**	-0,12	0,24	1		
$\log(\Delta U)$	0,12	0,03	0,18	-0,37**	-0,54***	1	
ΔI	-0,44***	0,14	-0,38**	0,21	0,61***	-0,56***	1

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

6.2 Yhteisintegroituvuus

Ennen lopullisen mallin spesifiointia, täytyi muuttujien mahdollinen yhteisintegroatio tutkia. Jos muuttujien välillä vallitsee yhteisintegroatio, mutta sitä ei kontrolloida, johtaa se väärin spesifioituun malliin. Muuttujien välinen yhteisintegroatio selvitettiin ainoastaan endogeenisiksi oletetuille muuttujille eli asuntomarkkinamuuttujille sekä Google-indekseille. Yksikköjuuritarkastelun (taulukko 4) yhteydessä näiden kaikkien muuttujien havaittiin olevan I(1)-prosesseja, joten muuttujien yhteisintegroituvuutta voitiin testata kaikki kolme muuttujaa sisältävässä VECM-systeemissä. Yhteisintegroatiotestinä käytettiin Johansenin λ_{trace} -testiä. Testin tulokset ovat esitettyinä taulukossa 6.

TAULUKKO 6 Johansenin yhteisintegroatiotestin tulokset

Muuttujat	Johansenin λ_{trace} -testisuure		
	$r \geq 2$	$r \geq 1$	$r = 0$
$\log(\Delta AK)$, $\log(\Delta AH)$ ja $\log(\Delta G1)$	3,77	12,58	34,88
$\log(\Delta AK)$, $\log(\Delta AH)$ ja $\log(\Delta G2)$	7,69	25,12	58,49***

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

Taulukon 6 tulosten perusteella Google1-indeksin ja asuntomarkkinamuuttujien välillä ei ole yhteisintegroatiota. Google2-indeksin ja asuntomarkkinamuuttujien väliltä sen sijaan löydetään yksi yhteisintegroatiovektori. Yhteisintegroatiovektorin havaitsemiseen Google2-indeksin ja asuntomarkkinamuuttujien välillä suhtaudutaan kuitenkin varauksella, sillä yhteisintegroatiovektorin havaitseminen on todennäköisesti seurausta Google2-indeksin stationaarisuudesta jo tasomuuttujana. Jos yksikin VECM-systeemissä testattavista muuttujista ei ole todellisuudessa I(1)-prosessi, johtaa se lähes poikkeuksetta yhteisintegroatiovektorin havaitsemiseen. Tätä tulkintaa tukee aiemmin yksikköjuuritestien tulosten esittelyn yhteydessä todettu Google2-indeksin stationaarisuus 1 % merkitsevyytasolla, kun ADF-testi tehtiin ainoastaan vakiokomponentti huomioiden. Lisäksi on todettava, että asuntomarkkinamuuttujien ja Google2-indeksin välistä yhteisintegroatiota ei voida perustella myöskään aiemman kirjallisuuden tuloksilla, sillä yksikään aiemmista asuntomarkkinamuuttujien ja Google-indeksien välistä suhdetta kuvanneista tutkimuksista ei ole käyttänyt lopullisessa mallinnuksessa virheenkorjausmallia. Näillä perusteiden yhteisintegroatiovektorin havaitseminen tulkittiinkin virheelliseksi, minkä vuoksi lopulliseen malliin ei sisällytetty virheenkorjaus-termiä.

6.3 Granger-kausaisuus

Differoitujen asuntomarkkinamuuttujien ja Google-indeksien välisiä vaikutussuhteita analysoitiin Granger-kausaisuustestien avulla. Granger-kausaisuustestit tehtiin parittain, eli toisin sanoen analyysissä käytettiin ainoastaan kahdesta muuttujasta muodostettuja VAR-systeemejä.⁸⁰ Viivepituudet Granger-kausaisuustesteihin määritettiin Bayesilaisella informaatiokriteerillä.⁸¹ Robustien tulkintojen takaamiseksi testit tehtiin myös käyttämällä viivepituutena koko edellistä vuotta eli neljää edellistä vuosineljänneistä.⁸² Granger-kausaisuustestien tulokset ovat esitettyinä taulukossa 7 asuntomarkkinamuuttujille sekä Google-indekseille. Tulokset on jaettu erillisiin paneeleihin Google-indeksien mukaan.

Taulukon 7 tuloksista havaitaan kirjallisuudessaakin usein havaittu asuntokauppojen ja asuntohintojen edeltäjä-seuraajasuhde (ks. esim. Hohenstatt ym. 2011). Asuntokauppojen lukumäärän muutosten havaitaan Granger aiheuttavan asuntohinnoissa tapahtuneita muutoksia 10 % merkitsevyystasolla, mutta toiseen suuntaan Granger-kausaisuutta ei havaita. Asuntokauppojen edellisten periodien muutoksilla voitaisiin siis mahdollisesti indikoida tämän hetken asuntohintojen muutosta. On kuitenkin todettava, että havaittu edeltäjä-seuraajasuhde on voimassa ainoastaan lyhyen viiveen VAR-malleille, sillä neljän viiveen VAR-mallilla tätä tulosta ei havaita.

Granger-kausaisuustestien tuloksista havaitaan myös, että kumpikaan Google-indekseistä ei Granger aiheuta asuntomarkkinamuuttujia. Granger-kausaisuustestien tulokset indikoivat ennemminkin Google-indeksien muutosten määräytyvän asuntomarkkinamuuttujien kehityksen perusteella. Neljän viiveen VAR-mallilla asuntokauppojen lukumäärän muutokset Granger aiheuttavat muutoksia Google2-indeksissä 5 % merkitsevyystasolla, kun taas asuntojen hintojen havaitaan puolestaan Granger aiheuttavan Google1-indeksiä 5 % merkitsevyystasolla. Google1-indeksin muutoksilla vaikuttaisi siis olevan vahvempi yhteys asuntojen hintakehitykseen ja Google2-indeksin muutoksilla puolestaan kauppojen lukumäärän kehitykseen.⁸³ On kuitenkin korostettava, että tulkinta on sensitiivinen VAR-mallissa käytetyille viivepituudelle. Lyhemmän viiveen VAR-malleilla asuntomarkkinamuuttujat eivät Granger-aiheuta Google-indeksejä.

⁸⁰ Granger-kausaisuustestien kuvailun yhteydessä käytiin läpi esimerkinomaisesti vastaava kahden muuttujan VAR-systeemi (kaavat 16 ja 17).

⁸¹ Joillekin systeemeistä ehdotettu viivepituus oli 0. Näille VAR-systeemeille käytettiin viivepituutena yhtä neljänneistä.

⁸² Pidemmän viivepituuden käyttöä tukee myös luvussa 2 esitetty asunnon ostoprosessikuvaus, jonka mukaan internetin käyttö asunnon etsinnän tietolähteenä voi kestää alle kahdesta kuukaudesta yli vuoteen.

⁸³ Luvun 2 ostoprosessikuvauksen perusteella parittaiset Granger-kausaisuustestit tehtiin myös molemmille Google-indekseille ja asuntomarkkinamuuttujille käyttämällä viivepituutena kahta ja kolmea neljänneistä. Näilläkin viivepituuksilla differoitujen Google-indeksien ei havaittu Granger aiheuttavan differoitujen asuntomarkkinamuuttujien arvoja.

TAULUKKO 7 Granger-kausalisuustestien tulokset

Paneeli A. Granger-kausalisuustestit asuntomarkkinamuuttujille ja Google1-indeksille			
Selittävät muuttujat	Selitettävät muuttujat		
	log(Δ AK)	log(Δ AH)	log(Δ G1)
F-testin arvot			
log(Δ AK)	-	2,91*(1)	1,05 (1)
log(Δ AH)	1,08 (1)	-	0,09 (1)
log(Δ G1)	0,18 (1)	0,25 (1)	-
log(Δ AK)	-	0,70 (4)	0,80 (4)
log(Δ AH)	0,13 (4)	-	3,24**(4)
log(Δ G1)	0,17 (4)	0,61 (4)	-
Paneeli B. Granger-kausalisuustestit asuntomarkkinamuuttujille ja Google2-indeksille			
Selittävät muuttujat	Selitettävät muuttujat		
	log(Δ AK)	log(Δ AH)	log(Δ G2)
F-testin arvot			
log(Δ AK)	-	2,91*(1)	2,33 (2)
log(Δ AH)	1,08 (1)	-	1,00 (2)
log(Δ G2)	0,47 (2)	0,71 (2)	-
log(Δ AK)	-	0,70 (4)	2,85**(4)
log(Δ AH)	0,13 (4)	-	0,86 (4)
log(Δ G2)	0,24 (4)	0,56 (4)	-

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. VAR-mallissa käytetty viivepituus on esitetty suluissa. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

6.4 Kontemporaalinen tarkastelu

Granger-kausalisuustestien tuloksista havaittiin, että differoitujen Google-indeksien viiveet eivät sisällä informaatiota asuntomarkkinamuuttujien muutoksista. On kuitenkin mahdollista, että kontemporaalisten Google-indeksien avulla pystytäisiin selittämään asuntomarkkinamuuttujien muutoksia.

Kontemporaalisten Google-indeksien kykyä selittää asuntomarkkinamuuttujien muutoksia tutkittiin kahdella yksinkertaisella aikasarjamallilla. Mallissa 24

differoidun asuntomarkkinamuuttujan nykyistä arvoa selitetään sen edellisen neljänneksen arvolla.⁸⁴ Mallissa 25 tähän malliin lisätään selittäväksi muuttujaksi differoitu kontemporaalinen Google-indeksi.

$$\text{Malli(24): } \log(y_t) = \alpha_t + \log(y_{t-1}) + u_t \quad (24)$$

$$\text{Malli(25): } \log(y_t) = \alpha_t + \log(y_{t-1}) + \log(x_t) + u_t \quad (25)$$

jossa α_t on vakiotermi, y_t viittaa joko asuntokauppojen lukumäärään tai asuntojen hintoihin, x_t viittaa puolestaan Google-indeksiin ja u_t on mallin virhetermi.

Taulukossa 8 ovat esitettyinä malleilla 24 ja 25 saadut tulokset asuntokaupoille sekä kontemporaaliselle Google1-indeksille. Tuloksista havaitaan molempien testattavien mallien aineistoon sopivuuden olevan erittäin heikko sopeutetulla R^2 -suureella tarkasteltuna. Kontemporaalisen Google1-indeksin muutosten ei havaita selittävän tilastollisesti merkitsevästi muutoksia asuntokauppojen lukumäärässä. Mallin aineistoon sopivuus ei myöskään merkittävästi parane, kun Google1-indeksi lisätään malliin. Kokonaisuutena voidaankin todeta, että kontemporaalisen Google1-indeksin muutokset eivät sisällä merkittävää asuntokauppojen lukumäärän muutoksia selittävää informaatiota.⁸⁵ Taulukon 8 mukaisia tuloksia ei raportoida lainkaan Google2-indeksille ja asuntokaupoille, sillä kontemporaalisen Google2-indeksin merkitys asuntokauppojen lukumäärän kehityksen selittäjänä havaittiin merkityksettömäksi.

TAULUKKO 8 Asuntokauppojen lukumäärä ja kontemporaalinen Google1-indeksi

Malli	(24)	(25)
Selittävät muuttujat		
$\log(\Delta AK)_{t-1}$	-0,07 (0,16)	-0,02 (0,15)
$\log(\Delta G1)$		0,23 (0,14)
Vakiotermi	-0,01 (0,01)	-0,01 (0,01)
Sopeutettu R^2	-0,02	0,02
AIC	-2,04	-2,06
BIC	-1,96	-1,93

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä estimaattien keskivirheet. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

⁸⁴ Muuttujien optimaalista viivepituutta arvioitiin Bayesilaisella informaatiokriteerillä. Optimaalinen viivepituus asuntokauppojen AR-malliin oli nolla ja asuntohinnoille yksi.

⁸⁵ Makrotaloudellisista muuttujista ainoastaan differoitu kontemporaalinen asuntolainan korko on tilastollisesti merkitsevä selittäjä asuntokauppojen lukumäärän kehitykselle (ks. liite 6 osio A). Makrotaloudelliset muuttujat sisältävä malli sopii myös aineistoon selvästi parhaiten.

Taulukko 9 sisältää mallien 24 ja 25 tulokset, kun selitettävänä muuttujana ovat kerrostaloasuntojen hintojen muutokset. Taulukosta 9 havaitaan differoidun Google1-indeksin olevan 10 % tasolla tilastollisesti merkitsevä selittäjä asuntohintojen muutoksille. Lisäksi aineistoon sopivuusmittareista sopeutettu R^2 paranee 5 prosenttiyksikköä, kun Google1-indeksi lisätään malliin. Jos aineistoon sopivuutta tarkastellaan informaatiokriteereillä, ei Google1-indeksin lisäämisen vaikutuksen tulkinta ole yksiselitteinen. Google1-indeksin lisääminen pienentää hieman Akaiken informaatiokriteeriä, mutta suurentaa puolestaan Bayesilaisen informaatiokriteerin arvoa verrattuna malliin 24. Kokonaisuutena muutos informaatiokriteerien arvoissa on kuitenkin hyvin vähäinen mallien 24 ja 25 välillä. Lisäksi on todettava, että Google1-indeksin tilastollinen merkitsevyys on sensitiivinen makrotaloudellisten muuttujien huomioimiselle. Kun makrotaloudelliset muuttujat lisätään malliin, Google1-indeksin tilastollinen merkitsevyys katoaa (ks. liite 6 osio B), mikä viittaisi makrotaloudellisten muuttujien sisältävän suurelta osin Google1-indeksin sisältämän informaation. Makrotaloudellisilla muuttujilla täydennetyssä mallissa tilastollisesti merkitseviä muuttujia ovat asuntohintojen viiveen lisäksi käytettävissä olevat tulot henkeä kohti sekä työttömyysaste. Lisäksi makrotaloudellisilla muuttujilla täydennetty malli sopii aineistoon parhaiten sopeutetun R^2 -suureen sekä Akaiken informaatiokriteerin perusteella. Taulukon 9 mukaisia tuloksia ei raportoida asuntojen hinnoille ja kontemporaaliselle Google2-indeksille, sillä Google2-indeksi ei ollut tilastollisesti merkitsevä selittäjä myöskään asuntojen hinnoille.⁸⁶

TAULUKKO 9 Asuntojen hinnat ja kontemporaalinen Google1-indeksi

Malli	(24)	(25)
Selittävät muuttujat		
$\log(\Delta AH)_{t-1}$	0,47*** (0,14)	0,49*** (0,14)
$\log(\Delta G1)$		0,04* (0,02)
Vakiotermi	0,01* (0,02)	0,00* (0,00)
Sopeutettu R^2	0,20	0,25
AIC	-5,88	-5,91
BIC	-5,80	-5,79

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluisissa ovat esitettyinä estimaattien keskivirheet. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

⁸⁶ Myös kontemporaalisten asuntomarkkinamuuttujien arvojen merkitys Google-indeksien nykyhetken arvoille tutkittiin. Mallin 25 mukaisesti tehdyissä tarkasteluissa havaittiin, että kumpikaan kontemporaalisista asuntomarkkinamuuttujista ei selittänyt tilastollisesti merkitsevästi Google1- eikä Google2-indeksin nykyhetken arvoja.

6.5 Aineistoon sopivuus

Granger-kausaisuustestien perusteella havaittiin, että Google-indeksejä ei ehkä ole järkevää käsitellä endogeenisina asuntomarkkinamuuttujia mallintavassa VAR-mallissa vaan pikemminkin kontemporaalisina eksogeenisina muuttujina. Lyhyen viiveen VAR-mallilla saatujen Granger-kausaisuustestien tuloksista havaittiin myös, että heikon eksogeenisuuden ehto on voimassa asuntomarkkinamuuttujien ja Google-indeksien välillä, eli toisin sanoen differoidut asuntomarkkinamuuttujat eivät Granger aiheuta differoitujen Google-indeksien arvoja. Biersensin (2004) mukaan tämän ehdon on täytyttävä, jotta muuttuja voidaan lisätä eksogeenisena VAR-malliin.

Lopullista mallia spesifioitaessa Google2-indeksi jätettiin kokonaan tarkastelun ulkopuolelle, sillä Granger-kausaisuustesteissä ja kontemporaalisessa tarkastelussa havaittiin, että Google2-indeksin muutokset eivät sisällä oleellista informaatiota kummankaan asuntomarkkinamuuttujan muutoksista. Mallinnuksessa huomioitiin näin ollen ainoastaan Google1-indeksi, jota kohdellaan VAR-mallissa eksogeenisena kontemporaalisena muuttujana. Aineistoon sopivuusanalyysissä sovellettu VAR-malli on esitetty matriisimuotoisena kaavassa 26. Kaikki kaavan 26 muuttujat ovat asuntolainan korkoa lukuun ottamatta log-transformoituja.

$$\begin{pmatrix} \Delta AK_t \\ \Delta AH_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \Delta AK_{t-i} \\ \Delta AH_{t-i} \end{pmatrix} + \mathbf{B} \begin{pmatrix} \Delta Y_t \\ \Delta U_t \\ \Delta I_t \\ \Delta G_t \end{pmatrix} + \begin{pmatrix} u_{AKt} \\ u_{Aht} \end{pmatrix} \quad (26)$$

jossa i kuvaa mallin viiveitä, ΔAK_t asuntokauppojen lukumäärä ΔAH_t asuntojen hintoja. Lisäksi eksogeenisina VAR-mallissa kohdellaan luvussa viisi esiteltyjä makrotaloudellisia muuttujia, ΔY_t , ΔU_t ja ΔI_t , sekä Google1-indeksiä, ΔG_t . \mathbf{A} on puolestaan endogeenisten muuttujien viiveiden ja \mathbf{B} eksogeenisten muuttujien kerroinmatriisi. u kuvaa kunkin yhtälön virhetermiä.

Taulukossa 10 ovat esitettyinä tulokset edellä kaavassa 26 kuvatun VAR-mallin aineistoon sopivuudelle. Aineistoon sopivuutta tarkasteltiin sopeutetulla R^2 -suureella sekä kahdella informaatiokriteerillä, BIC ja AIC. VAR-mallin aineistoon sopivuutta analysoitiin rajoittamalla kaavan 26 mukaiseen VAR-systeemiin sisällytettyjen eksogeenisten muuttujien lukumäärää. VAR-systeemissä olivat siis aina endogeenisiksi tulkitut asuntomarkkinamuuttujat, joiden lisäksi malliin sisällytettiin vaihteittain eri eksogeenisiä muuttujia. Näin pyrittiin saamaan selville Google1-indeksin merkitys asuntomarkkinamuuttujia kuvaavan mallin aineistoon sopivuudelle. VAR-mallin viivepituutena käytettiin yhtä neljänestä, joka valittiin Bayesilaisella informaatiokriteerillä.

Taulukon 10 paneelista A havaitaan, että paras asuntokauppojen lukumäärän muutosta kuvaava VAR-malli sisältää endogeenisten asuntohintojen lisäksi ainoastaan eksogeeniset makromuuttujat. Jos Google1-indeksi on ainoana eksogeenisena muuttujana, mallin aineistoon sopivuus on erittäin heikko. Kun Google1-indeksi sisällytetään malliin yhdessä makrotaloudellisten muuttujien

kanssa, havaitaan mallin aineistoon sopivuuden heikkenevän verrattuna malliin, jossa makrotaloudelliset muuttujat ovat ainoita eksogeenisiä muuttujia. Tulos Google1-indeksin heikosta asuntokauppojen lukumäärän indikointikyvystä ei ole yllättävä, sillä kontemporaalinen Google1-indeksi ei ollut tilastollisesti merkitsevä asuntokauppojen lukumäärää mallintavassa yksinkertaisessa mallissakaan.

Kuten taulukon 9 kontemporaalisessa tarkastelussa havaittiin, differoidun Google1-indeksin merkitys on suurempi asuntohintojen muutosten selittäjänä. Taulukon 10 paneelista B nähdään, että eksogeenisistä muuttujista pelkät makrotaloudelliset muuttujat sisältävän mallin aineistoon sopivuus on parempi kuin pelkän Google1-indeksin sisältävän mallin. Paras asuntojen hintojen muutosta kuvaava VAR-malli sisältää sopeutetun R^2 -suureen mukaan endogeenisen asuntokauppojen lukumäärän lisäksi kaikki eksogeeniset muuttujat, niin makrotaloudelliset muuttujat kuin Google1-indeksinkin. Informaatiokriteereillä tarkasteltuna paras asuntojen hintoja kuvaava malli sisältää kuitenkin eksogeenisistä muuttujista ainoastaan makrotaloudelliset muuttujat. Kokonaisuutena voidaan todeta, että kontemporaalinen Google1-indeksi tuo informaatiota asuntohinnoissa tapahtuvista muutoksista, mutta sen sisältämä informaatio on jossain määrin samankaltaista makrotaloudellisten muuttujien sisältämän informaation kanssa.

TAULUKKO 10 VAR-mallin aineistoon sopivuus

Paneeli A: Asuntokauppojen lukumäärää selittävän mallin aineistoon sopivuus			
Sopeutettu R^2	0,24	0,03	0,22
Akaike, AIC	-2,25	-2,05	-2,21
Schwarz, BIC	-2,00	-1,88	-1,92
Google1-indeksi		x	x
Makromuuttujat	x		x
Paneeli B: Asuntojen hintoja selittävän mallin aineistoon sopivuus			
Sopeutettu R^2	0,31	0,24	0,34
Akaike AIC	-5,98	-5,88	-5,97
Schwarz BIC	-5,81	-5,71	-5,67
Google1-indeksi		x	x
Makromuuttujat	x		x

6.6 Aineiston ulkopuolinen ennustekyky

Aineistoon sopivuustarkasteluissa kontemporaalisen Google1-indeksin muutosten havaittiin indikoivan jossain määrin kerrostaloasuntojen hintojen muutoksia, kun taas asuntokauppojen lukumäärän muutoksista Google1-indeksi sisälsi vähemmän informaatiota. Aineistoon sopivuuden analysoinnin jälkeen muuttujista muodostettiin ennustemalleja, joiden avulla pyritään selvittämään, voidaanko lähitulevaisuudessa tapahtuvia asuntomarkkinoiden muutoksia ennustaa tarkemmin Google1-indeksin avulla. Molemmille asuntomarkkinamuuttujille laadittiin yhden askeleen ennusteet käyttämällä rekursiivista ennustusmenetelmää. Aiemmassa Google-hakuaineistoa asuntomarkkinoiden indikaattorina tutkivassa kirjallisuudessa muun muassa Dietzel ym. (2014) on soveltanut ennustetarkasteluissa vastaavaa ennustusmenetelmää.

Rekursiivisia ennusteita laadittaessa VAR-mallin parametrit estimoitiin ensimmäisen kerran aikavälillä 1:2004-4:2010. Tällä aineistolla estimoidulla VAR-mallilla ennustettiin seuraavan neljänneksen havainnon arvoa eli vuoden 2011 ensimmäisen neljänneksen arvoa. Tämän ennusteen estimoinnin jälkeen ennustettu havainto, 1:2011, sisällytettiin ennustamiseen käytettävään aineistoon, minkä jälkeen VAR-mallin parametrit estimoitiin uudelleen käyttämällä estimointiaineistona aikaväliä 1:2004-1:2011. Tällä aikavälillä estimoidulla VAR-mallilla ennustettiin jälleen yksi askel eteenpäin laatien ennuste vuoden 2011 toiselle neljännekselle, minkä jälkeen tämä havainto lisättiin parametrien estimointiin käytettävään aineistoon. Yhden askeleen ennustamiseen käytettyä aineistoa kasvatettiin edellä kuvatulla tavalla aina vuoden 2014 toiseen neljännekseen saakka. Viimeinen ennuste laadittiin käyttämällä VAR-mallin estimointiaineistona aikaväliä 1:2004-2:2014, ja ennustamalla sillä ajanhetken 3:2014 arvoa.⁸⁷

Edellä kuvatut rekursiiviset yhden askeleen ennusteet estimoitiin neljällä eri VAR-mallilla. Käytetyt VAR-mallit ovat esitettyinä taulukossa 11 sekä kaavoissa 27-30. Asuntolainan korkoa lukuun ottamatta kaavojen 27-30 kaikki muuttujat ovat log-transformoituja. Bayesilaisen informaatiokriteerin perusteella ennustemallit estimoitiin yhden viiveen VAR-malleina.

TAULUKKO 11 Ennustemallit

Malli	Ennustettava muuttuja	Kaupat/Hinnat	Google1-indeksi	Makromuuttujat
m1	Kaupat/hinnat	x		
g1	Kaupat/hinnat	x	x	
m2	Kaupat/hinnat	x		x
g2	Kaupat/hinnat	x	x	x

⁸⁷ Ennustettavia havaintoja oli yhteensä 15.

$$\text{Malli}(m1): \begin{pmatrix} \Delta AK_t \\ \Delta AH_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \Delta AK_{t-i} \\ \Delta AH_{t-i} \end{pmatrix} + \begin{pmatrix} u_{AKt} \\ u_{AHt} \end{pmatrix} \quad (27)$$

$$\text{Malli}(g1): \begin{pmatrix} \Delta AK_t \\ \Delta AH_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \Delta AK_{t-i} \\ \Delta AH_{t-i} \end{pmatrix} + \mathbf{B}(\Delta G_t) + \begin{pmatrix} u_{AKt} \\ u_{AHt} \end{pmatrix} \quad (28)$$

$$\text{Malli}(m2): \begin{pmatrix} \Delta AK_t \\ \Delta AH_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \Delta AK_{t-i} \\ \Delta AH_{t-i} \end{pmatrix} + \mathbf{B} \begin{pmatrix} \Delta Y_t \\ \Delta U_t \\ \Delta I_t \end{pmatrix} + \begin{pmatrix} u_{AKt} \\ u_{AHt} \end{pmatrix} \quad (29)$$

$$\text{Malli}(g2): \begin{pmatrix} \Delta AK_t \\ \Delta AH_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \Delta AK_{t-i} \\ \Delta AH_{t-i} \end{pmatrix} + \mathbf{B} \begin{pmatrix} \Delta Y_t \\ \Delta U_t \\ \Delta I_t \\ \Delta G_t \end{pmatrix} + \begin{pmatrix} u_{AKt} \\ u_{AHt} \end{pmatrix} \quad (30)$$

jossa i kuvaa jälleen mallin viiveitä, ΔAK_t asuntokauppoja, ΔAH_t asuntojen hintoja. Eksogeenisina VAR-mallissa kohdellaan luvussa viisi esiteltyjä makrotaloudellisia muuttujia, ΔY_t , ΔU_t ja ΔI_t , sekä Google1-indeksiä, ΔG_t . \mathbf{A} on puolestaan endogeenisten muuttujien viiveiden ja \mathbf{B} kontemporaalisten eksogeenisten muuttujien kerroinmatriisi. u kuvaa kunkin yhtälön virhetermiä.

Taulukossa 12 ovat esitettyinä valittujen ennustetarkkuusmittareiden arvot analysoiduille ennustemalleille m1, m2, g1 ja g2. Paneeli A käsittää ennustetarkasteluiden tulokset asuntokauppojen lukumäärälle ja paneeli B puolestaan asuntojen hinnoille. Ennusteiden tarkkuutta mitataan luvussa neljä käsitellyillä keskineliövirheellä (MSE) sekä Theilin U1-kertoimella (U1).⁸⁸ Asuntomarkkinamuuttujien kehitykselle estimoitujen yhden askeleen ennusteiden ja toteutuneiden arvojen suhdetta on lisäksi havainnollistettu graafisesti kuvioissa 6 ja 7.

Taulukon 12 paneelin A tuloksista havaitaan, että eksogeenisten makrotaloudellisten muuttujien lisääminen asuntokauppojen lukumäärän muutosten ennustemalliin johtaa merkittävimpään parannukseen ennustetarkkuudessa. Google1-indeksin sisältävät g1- ja g2-mallit eivät erotu ennustetarkkuudeltaan vertailtavista perusmalleista, m1- ja m2-malli. Google1-indeksin sisältämä g1-malli on U1-kertoimen avulla tarkasteltuna ainoastaan 1,2 prosenttia parempi verrattuna ainoastaan asuntomarkkinamuuttujat sisältävään m1-malliin. Eksogeeniset makrotaloudelliset muuttujat sisältäviä ennustemalleja, m2 ja g2, tarkasteltaessa havaitaan Google1-indeksin lisäämisen vaikutuksen olevan merkityksetön mallin ennustetarkkuudelle. Theilin U1-kertoimen arvon perusteella voidaan myös todeta, että asuntokauppojen lukumäärän ennusteet ovat kaikilla analysoiduilla malleilla melko epätarkkoja.

⁸⁸ Ennustetarkkuuden tarkasteluissa keskitytään ensisijaisesti Theilin U1-kertoimen tulkin-taan, sillä keskineliövirheen arvot jäävät erittäin pieniksi. Ennustetarkastelut suoritettiin myös tekemällä muuttujille anti-log-muunnos, mutta keskineliövirheet olivat myös täl-löin yhtä pieniä.

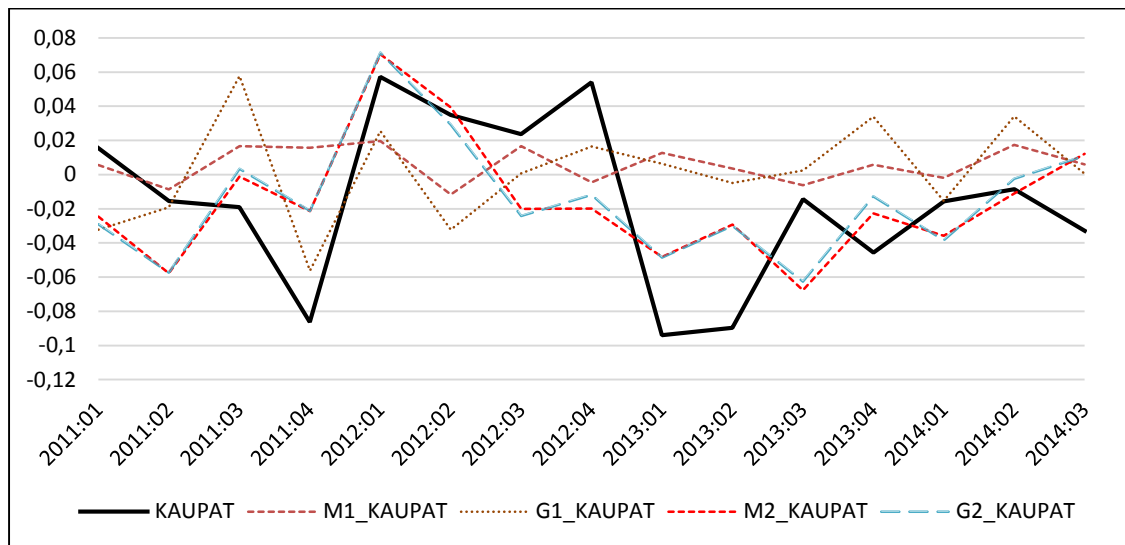
Ennustemallien tarkkuudet asuntojen hintojen muutoksille raportoidaan puolestaan taulukon 12 B paneelissa. Asuntokauppojen lukumäärän ennusteista tehdyt tulkinnat ovat voimassa pääsääntöisesti myös asuntohintojen ennustetarkastelussa. Asuntohintojen muutoksia ennustavien mallien ennustetarkkuuden havaitaan myös paranevan eniten, kun ennustemalliin lisätään eksogeeniset makrotaloudelliset muuttujat. Yksinkertaisemmilla malleilla, m1 ja g1, tarkasteltuna Google1-indeksin lisääminen pienentää Theilin U1-kertoimen arvoa 1,1 prosenttia, mutta rikkaampien mallien, m2 ja g2, ennustetarkkuutta vertailtaessa Google1-indeksin lisääminen malliin ei vaikuta ennustetarkkuuteen.

Mielenkiintoiseksi tulokseksi voidaan tulkita se, että Theilin U1-kerroin on suurempi asuntohintojen ennustemallille verrattuna asuntokauppojen lukumäärän ennustemalliin. Toisin sanoen estimoidut ennusteet ovat siis tarkempia asuntokauppojen lukumäärän muutoksille (vrt. esim. taulukko 10). Tosin on korostettava, että kummallekaan asuntomarkkinamuuttujalle ennusteet eivät ole erityisen tarkkoja.⁸⁹ Lisäksi makrotaloudellisten muuttujien lisääminen ennustemalleihin aiheuttaa suuremman ennustetarkkuuden parannuksen asuntokauppojen lukumäärän ennusteissa, mikä indikoi makrotaloudellisten muuttujien sisältävän enemmän informaatiota asuntokauppojen lukumäärän kuin asuntojen hintojen kehityksestä. Google1-indeksi ei puolestaan vaikuttaisi tuovan ennustetarkkuutta oleellisesti parantavaa informaatiota kummankaan asuntomarkkinamuuttujan muutoksista.

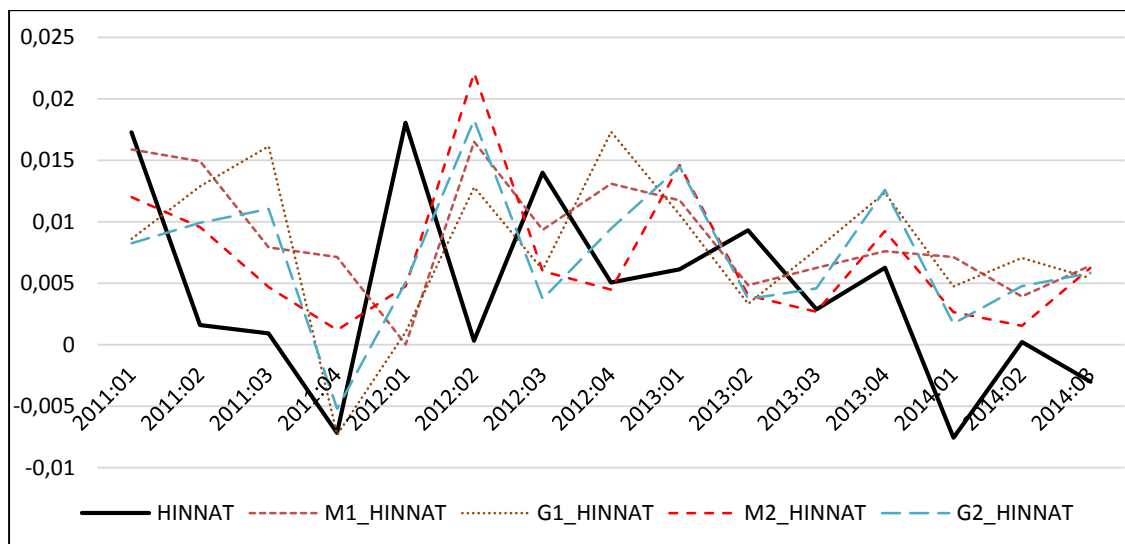
TAULUKKO 12 Yhden askeleen ennusteet asuntomarkkinamuuttujille

Paneeli A. Yhden askeleen ennustetarkkuus asuntokauppojen lukumäärälle			
Malli	Ennustettava muuttuja	MSE	U1
m1	Asuntokaupat	0,003	0,85
g1	Asuntokaupat	0,003	0,84
m2	Asuntokaupat	0,002	0,66
g2	Asuntokaupat	0,002	0,66
Paneeli B. Yhden askeleen ennustetarkkuus asuntojen hinnoille			
Malli	Ennustettava muuttuja	MSE	U1
m1	Asuntojen hinnat	0,0001	0,89
g1	Asuntojen hinnat	0,0001	0,88
m2	Asuntojen hinnat	0,0001	0,79
g2	Asuntojen hinnat	0,0001	0,79

⁸⁹ Esimerkiksi Dietzel ym. (2014) tutkimuksessa liikekiinteistöjen kauppalukumäärien ja hintojen ennusteiden Theilin U1-kerroin vaihtelee 0,052 ja 0,210 välillä.



KUVIO 6 Asuntokauppojen lukumäärä ja ennustemallien rekursiiviset yhden askeleen ennusteet



KUVIO 7 Asuntojen hinnat ja ennustemallien rekursiiviset yhden askeleen ennusteet

6.7 Tulosten robustisuus

Tulosten robustisuutta tutkittiin analysoimalla vaihtoehtoisien kausitasoitustemelmän sekä otoskoon merkitystä. Tämän lisäksi tulosten robustisuutta tarkasteltiin vielä muuntamalla aineiston nimelliset muuttujat reaalisiksi. Muunnosten merkitystä kartoitettiin Granger-kausalisuustesteillä sekä analysoimalla luvussa 6.4 esitetyillä malleilla kontemporaalisen Google-indeksin kykyä selittää asunto-markkinamuuttujien liikkeitä.

Edellä esitetyissä tuloksissa kausitasoitus suoritettiin X11 Census-menetelmällä. Tulosten sensitiivisyyttä kausitasoitusmenetelmän valinnalle analysoitiin kausitasoittamalla muuttujat X11 Census-menetelmän sijasta ottamalla muuttujista vuosidifferenssi eli vähentämällä tämän neljänneksen arvosta edellisen vuoden vastaavan neljänneksen arvo.⁹⁰ Vuosidifferoiduista muuttujista asuntomarkkinamuuttujat olivat stationaarisia, kun taas Google-indeksien havaittiin olevan epästationaarisia. Kaikista tarkastelluista muuttujista otettiin kuitenkin vielä yksi differenssi stationaarisuuden takaamiseksi.⁹¹

Vuosidifferoitujen muuttujien välisiä suhteita analysoitiin aluksi parittaisten Granger-kausaisuustestien avulla (ks. liite 7). Kausaisuustestien tulokset eivät juurikaan poikenneet aiemmin taulukossa 7 esitetyistä tuloksista. Asuntokauppojen lukumäärän havaitaan nytkin Granger aiheuttavan asuntohintoja yhden viiveen VAR-mallilla. Tämä tulos häviää kuitenkin jälleen neljän viiveen VAR-mallia käytettäessä. Myös asuntomarkkinamuuttujien ja Google-indeksien välisen suhteen havaitaan olevan samankaltainen taulukon 7 tulosten kanssa. Neljän viiveen VAR-malleilla asuntohintojen muutosten havaitaan Granger aiheuttavan Google1-indeksin muutoksia 1 % merkitsevyytasolla, kun taas asuntokauppojen lukumäärän muutosten havaitaan Granger aiheuttavan Google2-indeksin muutoksia 1 % merkitsevyytasolla. Informaation ei havaita myöskään vuosidifferoiduille muuttujilla kulkevan Google-indekseistä asuntomarkkinamuuttujiin.

Mallien 24 ja 25 kuvaamalla tavalla analysoitiin myös vuosidifferoitujen kontemporaalisten Google-indeksien kykyä selittää vuosidifferoitujen asuntomarkkinamuuttujien arvoja. Google1-indeksin havaittiin olevan tilastollisesti merkitsevä selittäjä molemmille asuntomarkkinamuuttujille, kun taas Google2-indeksin kontemporaaliset arvot eivät kyenneet kuvaamaan kummankaan asuntomarkkinamuuttujien nykyhetken arvoja. Kontemporaalinen Google1-indeksi oli tilastollisesti 5 % tasolla merkitsevä selittävä tekijä molemmille asuntomarkkinamuuttujille (ks. liite 8). Vuosidifferoiduilla muuttujilla kontemporaalisen Google1-indeksin havaittiin siis selittävän myös asuntokauppojen lukumäärää. X11 Census-menetelmällä kausitasoitetuilla muuttujilla tarkasteltuna kontemporaalisen Google1-indeksin muutosten ei havaittu selittävän kuin ainoastaan muutoksia asuntojen hinnoissa ja niitäkin ainoastaan 10 % merkitsevyytasolla. Tosin on korostettava, että myös vuosidifferoiduilla muuttujilla Google1-indeksin tilastollinen merkitsevyys katoaa, kun malliin 25 lisätään makrotaloudelliset muuttujat.

Kausitasoitusmenetelmän lisäksi tulosten robustisuutta tutkittiin käyttämällä eripituista otosta. Asuntomarkkinamuuttujien ja Google-indeksien suhteen sensitiivisyyttä otoksen valinnalle analysoitiin tekemällä parittaiset Granger-kausaisuustestit ja kontemporaalinen tarkastelu aikavälillä 1:2004–1:2008. Aikavälin

⁹⁰ Aiemmassa kirjallisuudessa esimerkiksi Dietzel ym. (2014) ovat käyttäneet kausitasoitukseen tätä menetelmää.

⁹¹Jo valmiiksi stationaaristen asuntomarkkinamuuttujien kohdalla yhden ylimääräisen differenssin ottaminen saattaa johtaa ylidifferointiin ja sitä kautta merkityksellisen informaation katoamiseen. Tämän vuoksi Granger-kausaisuustestit tehtiin myös VAR-systeemeillä, joissa ainoastaan vuosidifferoiduista ja epästationaarisista Google-indekseistä otettiin differenssi. Granger-kausaisuustestien tulokset eivät kuitenkaan olleet sensitiivisiä tämän ylimääräisen differenssin ottamiselle, minkä vuoksi päädyttiin ottamaan kaikista muuttujista differenssi.

viimeiseksi havainnoksi valittiin vuoden 2008 ensimmäinen neljännes Zivot-Andrews-testin tulosten perusteella (ks. liite 4). Pienemmän otoskoon tarkastelussa muuttajat kausitasoitettiin X11 Census-menetelmällä ja yksikköjuuritestien tulosten perusteella muuttujista käytettiin differenssimuuttujia.

Pienemmällä otoksella saatujen Granger-kausalisuustestien tulosten mukaan Google-indeksit Granger aiheuttavat asuntokauppojen lukumäärän arvoja (ks. liite 9), mutta kumpikaan Google-indekseistä ei Granger aiheuta asuntohintamuuttujan arvoja. Google2-indeksi Granger aiheuttaa asuntokauppojen lukumäärän nykyhetken arvon 1 % merkitsevyytasolla ja Google1-indeksi puolestaan 10 % merkitsevyytasolla. Kahdensuuntaista Granger-kausalityyttä ei ole havaittavissa käytetyllä pienemmällä otoksella, sillä asuntokauppojen lukumäärä ei Granger aiheuta Google-indeksien nykyhetken arvoja. Nämä pienemmällä otoksella saadut Granger-kausalisuustestien tulokset tukevat kuviosta 4 tehtyjä päätelmiä muuttajien edeltäjä-seuraaja suhteesta ennen finanssikriisin alkua. Pienemmällä otoksella tutkittiin myös mallien 24 ja 25 mukaisesti kontemporaalisten Google-indeksien merkitystä asuntomarkkinamuuttujien nykyhetken arvoille. Kontemporaalistet Google-indeksit eivät kuitenkaan olleet tilastollisesti merkitseviä selittäviä tekijöitä kummallekaan asuntomarkkinamuuttujalle.

Viimeisenä robustisuustarkasteluna analysoitiin aineistoa, jossa aineiston nimelliset muuttajat muunnettiin reaalisiksi Tilastokeskuksen kuluttajahintaindeksin avulla.⁹² Nimellisiä muuttujia tarkasteltaessa Google1-indeksin merkitys havaittiin suurimmaksi asuntojen hinnoille. Tämän vuoksi onkin syytä selvittää, muuttuko muuttajien suhde, kun asuntojen hinnoista käytetään nimellisten arvojen sijasta reaalisia arvoja. Reaalisille muuttujille tehtiin kausitasoitus myös X11 Census-menetelmällä ja yksikköjuuritestien tulosten perusteella muuttujista käytettiin differenssimuuttujia.

Parittaisista Granger-kausalisuustesteistä havaittiin, että myöskään Google-indeksien ja reaalisten asuntohintojen välillä ei ollut Granger kausalisuutta.⁹³ Erona taulukon 7 tuloksiin oli se, että reaaliset asuntojen hinnat eivät Granger aiheuttaneet Google1-indeksin arvoja millään testatulla viivepituudella. Kontemporaalisen Google1-indeksin ja reaalisten asuntojen hintojen välistä suhdetta tutkittiin myös mallien 24 ja 25 avulla. Kontemporaalisen Google-indeksin havaittiin selittävän reaalisten asuntohintojen arvoja 5 % merkitsevyytasolla (ks. liite 10). Toisin kuin nimellisiä muuttujia tarkasteltaessa Google1-indeksin tilastollinen merkitsevyys säilyy, vaikka makrotaloudelliset muuttujat lisätään malliin. Vaikka kontemporaalisen Google1-indeksin havaittiin selittävän tilastollisesti merkittävästi reaalisia asuntojen hintoja, kulkee informaatio myös toiseen suuntaan, sillä kontemporaalisten reaalisten asuntojen hintojen havaittiin selittävän myös

⁹² Aineiston muuttujista muunnos reaalisesta nimelliseksi vaikutti asuntojen hintoihin, käytettävissä oleviin tuloihin henkeä kohti sekä asuntolainan korkoon.

⁹³ Granger-kausalisuustestit tehtiin käyttämällä viivepituuksina yhtä neljänestä sekä neljää neljänestä. Bayesilainen informaatiokriteeri ehdotti kaikille malleille viivepituudeksi nollaa tai yhtä viivettä.

Google1-indeksin nykyhetken arvoa tilastollisesti merkitsevästi 5 % merkitsevyystasolla. ⁹⁴ Päättelmänä voidaan todeta, että kontemporaalisen Google1-indeksin merkitys asuntohintamuutosten selittäjänä vaikuttaisi olevan suurempi, kun selittävänä muuttujana ovat reaaliset asuntojen hinnat.

⁹⁴ Google2-indeksin kontemporaalisten arvojen havaittiin olevan tilastollisesti merkitsevä 10 % merkitsevyystasolla, kun makromuuttujat lisättiin malliin 25. Jos mallissa ei ollut makromuuttujia, kontemporaalinen Google2-indeksi ei ollut tilastollisesti merkitsevä selittäjä reaalisten asuntohintojen nykyhetken arvolla.

7 YHTEENVETO

Tämä tutkimus on tiettävästi ensimmäinen Suomen asuntomarkkinoille Google-hakuaineistoa soveltava tutkimus. Suomessa Google on ylivoimaisesti käytetyin hakukone ja kansalaisista yli 85 prosenttia käyttää internetiä säännöllisesti. Tämän lisäksi Suomessa asuntomarkkinoiden tarjonta on suurelta osin siirtynyt verkkoon. Nämä syyt antavatkin olettaa, että suomalaisiakin asuntomarkkinoita voitaisiin indikoida Googlessa tehtyjen verkkohakujen pohjalta.

Tutkimuksen tulokset eroavat kuitenkin suurelta osin aiemman tutkimuksen tuloksista. Aiemmassa kirjallisuudessa Google-indeksien on havaittu indikoivan sekä asuntokauppojen lukumäärän (ks. esim. Hohenstatt ym. 2011) että asuntojen hintojen (ks. esim. McLaren & Shanbhogue 2011) kehitystä ja parantavan merkittävästi mallien aineistoon sopivuutta sekä ennustetarkkuutta. Tässä tutkimuksessa koko otoksen tasolla tarkasteltuna ainoastaan kontemporaalisen Google1-indeksin havaitaan selittävän asuntomarkkinoiden muutoksia tilastollisesti merkitsevästi, mutta tämäkin tulos on sensitiivinen makrotaloudellisten muuttujien huomioimiselle. Lisäksi nimellisillä ja X11 Census-menetelmällä kausitasoitetuilla muuttujilla tarkasteltuna Google1-indeksin merkitys mallin aineistoon sopivuudelle sekä ennustetarkkuudelle on melko vähäinen. Aiemmassa kirjallisuudessa myös Wu ja Brynjolfsson (2013) havaitsevat kontemporaalisen Google-indeksin selittävän asuntomarkkinoiden kehitystä Google-indeksin viiveellisiä arvoja enemmän, mutta he havaitsevat Google-indeksin merkityksen suureksi myös tehdyille ennusteille, mitä tässä tutkimuksessa ei havaita.

Asuntomarkkinoiden kehitystä Google-hakuaineiston avulla indikoineet tutkimukset ovat myös havainneet kahdensuuntaisen Granger-kausaliiteetin asuntomarkkinoiden ja käytettyjen Google-indeksien välillä (ks. Hohenstatt ym. 2011, Beracha & Wintoki 2013, Hohenstatt & Käsbauer 2014 ja Dietzel ym. 2014). Tämän tutkimuksen tulokset indikoivat, että edellisten periodien Google-indeksien muutoksilla ei kyettäisi ennakoimaan asuntomarkkinamuuttujien muutoksia. Koko otoksella saatujen tulosten mukaan kumpikaan Google-indekseistä ei Granger aiheuta asuntomarkkinamuuttujia millään testatulla viivepituudella. Sen sijaan Granger-kausaliisuustestien tulokset puoltavat ennemminkin tulkintaa, jonka mukaan informaatio siirtyy asuntomarkkinoilta Google-indeksien arvoihin. Poikkeuksena saatuihin Granger-kausaliisuustestien tuloksiin ovat lyhemmällä otoksella saadut tulokset, joiden mukaan molempien Google-indeksien muutokset Granger aiheuttavat muutoksia asuntokauppojen lukumäärässä. Google-hakuaineiston ja asuntomarkkinamuuttujien suhde vaikuttaisikin olevan jossain määrin riippuvainen otoskoosta. Tämä tulos on havaittu myös aiemmassa kirjallisuudessa. Muun muassa Barreira ym. (2013) korostavat, että erityisesti kestokulutushyödykkeitä mallinnettaessa Google-hakuaineiston indikaatiokyky vaihtelee merkittävästi valitun tarkasteluperiodin mukaan. Lisäksi he huomauttavat, että henkilöiden käyttäytyminen kestokulutushyödykkeiden ostoprosessissa ei ole ajassa vakio.

Tämän tutkimuksen ja aiemman kirjallisuuden tulosten eroavaisuuteen on useita mahdollista syytä. Tavat, joilla informaatiota etsitään hakukoneiden kautta,

eroavat maittain ja samaakin asiaa tarkoittavilla hauilla voi olla erilainen merkitys eri maissa (ks. esim. McLaren & Shanbhogue 2011). Lisäksi informaation etsintävaiheen kesto kestokulutushyödykkeitä tutkittaessa eroaa merkittävästi tietoa etsivän henkilön mukaan (Barreira ym. 2013). Suomalaisesta asunnon ostoprosessista ja erityisesti asuntojen etsintävaiheen kestosta löydettiin melko vähän tutkimuksellista tietoa, minkä vuoksi mallinnuksessa käytetyt viivepituudet määriteltiin pääasiassa informaatiokriteerin sekä aineiston frekvenssin perusteella. Tällä tavoin määritellyt viivepituudet voivat kuitenkin poiketa todellisesta asunnon etsintävaiheen kestosta.

On myös mahdollista, että muodostetut Google-indeksit kuvaavat asuntomarkkinoita ja erityisesti kerrostaloasuntojen markkinoita liian yleisellä tasolla. Haluttua asuntomarkkinoita indikoivaa käyttäytymistä ei tällöin kyetä tunnistamaan riittäväällä tarkkuudella Google-indeksin sisältämän runsaan kohinan vuoksi. Tämä ongelma aiheutuu osittain siitä, että asuntomarkkinoita kuvaavia spesifimpiä sanoja ei juurikaan ole tai sitten niillä haetaan erittäin epäsäännöllisesti. Tämä on puolestaan seurausta siitä, että suomalaiset asuntomarkkinat ovat poikkeuksellisen vahvasti keskittyneet asuntoportaaleihin, joiden sisällä henkilöiden käyttäytymistä on mahdotonta havaita Google-hakuaineiston avulla. Ongelma olisi mahdollisesti kyetty välttämään osittain, jos Googlen valmiit luokittelukategoriat olisivat käytössä myös Suomen maantieteellisellä alueella. Useat aiemmat tutkimukset ovat havainneet positiivisia tuloksia Google-aineiston kyvystä indikoida asuntomarkkinoita hyödyntämällä nimenomaan Googlen valmiita luokittelukategorioita (ks. esim. Hohenstatt ym. 2011 ja Wu & Brynjolfsson 2013).

Vaikka tässä tutkimuksessa analysoitujen Google-indeksien merkitys valittujen asuntomarkkinamuuttujien indikaattorina havaittiinkin melko vähäiseksi, on erittäin todennäköistä, että Google-indeksien avulla kyetään parantamaan asuntomarkkinamuuttujia kuvaavia malleja. Tulevaisuudessa Google-hakuaineistoa Suomen asuntomarkkinoille soveltavan tutkimuksen tulisikin selvittää, mitä muita hakusanoja asuntomarkkinoita kuvaavien Google-indeksien muodostamisessa voitaisiin hyödyntää ja voitaisiinko Google-indeksin selityskykyä parantaa soveltamalla jotain aiemman kirjallisuuden käyttämää tilastollista menetelmää (ks. esim. Carrière-Swallow ja Labbé 2013). Tämän lisäksi tutkimuksissa tulisi selvittää, kuinka Google-hakuaineiston indikaatiokyky vaihtelee eri taloudellisissa suhdanteissa.

LÄHTEET

- Acemoglu, D. & Scott, A. 1994. Consumer confidence and rational expectations: are agents' beliefs consistent with the theory? *Economic Journal* 104, 154-194.
- Adams, Z. & Füß, R. 2010. Macroeconomic determinants of international housing markets. *Journal of Housing Economics* 19, 38-50.
- Agnello, L. & Schucknecht, L. 2011. Booms and busts in housing markets: determinants and implications. *Journal of Housing Economics* 20, 171-190.
- Alma Median lehdistötiedote. 2004. Ostopäätöstä edeltävä tiedonhankinta asuntokaupassa 2004 -tutkimus: Internetin merkitys asuntokaupan tiedonlähteenä kasvanut selvästi. Haettu 10.05.2015 osoitteesta: <URL:<http://www.almamedia.fi/uutishuone/tiedotteet/2004/#ostopaato-sta-edeltava-tiedonhankinta-asuntokaupassa-2004--tutkimus-internetin-merkitys-asuntokaupan-tiedonlahteena-kasvanut-selvasti>>.
- Askatas, N. & Zimmermann, K. F. 2009. Google econometrics and unemployment forecasting. *Applied Economics Quarterly* 55 (2), 107-120.
- Askatas, N. & Zimmermann, K. F. 2014. Detecting mortgage delinquencies with Google Trends. Julkaisematon.
- Askatas, N. & Zimmermann, K. F. 2015. The internet as a data source for advancement in social sciences. *International Journal of Manpower* 36 (1), 2-12.
- Barber, M. B. & Odean, T. 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21, 785-818.
- Barreira, N., Godinho, P. & Melo, P. 2013. Nowcasting the unemployment rate and car sales in south-western Europe with Google Trends. *Netnomics* 14, 129-165.
- Beracha, E. & Wintoki, M. B. 2013. Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research* 30 (3), 283-312.
- Belessiotis, T. 1996. Consumer confidence and consumer spending in France. *European Commission Economic Papers* 116.
- Bierens, J. H. 2004. VAR models with exogenous variables. Haettu 26.10.2015 osoitteesta:<URL:http://grizzly.la.psu.edu/~hbierens/EasyRegTours/VAR_Tourfiles/VARX.PDF>.
- Bollen, J., Mao, H. & Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1-8.
- Brooks, C. 2008. *Introductory econometrics for finance*. Cambridge. Cambridge University Press.
- Carrière-Swallow, Y. & Labbé, F. 2013. Nowcasting with Google Trends in emerging market. *Journal of Forecasting* 32 (4), 289-298.
- Carroll, D. C., Jeffrey, C. F. & Wilcox, W. D. 1994. Does consumer sentiment forecast household spending? If so, why? *American Economic Review* 84, 1397-1408.
- Case, K. E. & Schiller, R. J. 1989. The Efficiency of the market for single-family homes. *American Economic Review* 79 (1), 125-37.

- Case, K.E. & Schiller, R.J. 2003. Is there a bubble in the housing market? *Brookings Papers on Economic Activity* 2, 299-362.
- Castle, J. L., Fawcett, N. W. & Hendry, D. F. 2009. Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review* 210 (1), 71-89.
- Choi, H. & Varian, H. R. 2009a. Predicting initial claims for unemployment benefits. Technical Report, Google.
- Choi, H. & Varian, H. R. 2009b. Predicting the present with Google Trends. Technical Report, Google.
- Choi, H. & Varian, H. R. 2011. Predicting the present with Google Trends. Presentation at SF Fed, March 18.
- Choi, H. & Varian, H. R. 2012. Predicting the present with Google Trends. *Economic Record* 88 (1), 2-9.
- Cotsomitis, A. J. & Kwan, C., C., A. 2006. Can consumer confidence forecast household spending? Evidence from the European Commission business and consumer surveys. *Southern Economic Journal* 72 (3), 597-610.
- Da, Z., Engelberg, J. & Gao, P. 2011. In search of attention. *The Journal of Finance* 66 (5), 1461-1499.
- Dickey, D. A. & Fuller, W. A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74 (366), 427-431.
- Dietzel, A. M., Braun, N. & Schäfers, W. 2014. Sentiment-based commercial real estate forecasting with Google search volume data. *Journal of Property Investment & Finance* 32 (6), 540-569.
- Engle, R.F. & Granger C.W.J. 1987. Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55, 251-76.
- Englund, P. & Ioannides, Y. M. 1997. House price dynamics: an international empirical perspective. *Journal of Housing Economics* 6, 119-136.
- Etuovi.com. 2015. Markkinapuntari: Asuntokaupan kehitys 10/2015. Tarkastelu tehtiin seuraavilla rajauksilla: käytetyt asunnot ja kohdemäärä aikavälillä 2004/01-2015/10. Haettu 1.11.2015 osoitteesta: <URL: <http://www.etuovi.com/market-barometer?locale=en>>.
- European Commission: Economic and Financial Affairs. Haettu 28.8.2015 osoitteesta:<URL:http://ec.europa.eu/economy_finance/db_indicators/surveys/time_series/index_en.htm>.
- Fondeur, Y. & Karamé, F. 2013. Can Google data help predict french youth employment? *Economic Modelling* 30, 117-125.
- Fu, Y. & Ng, K. L. 2001. Market efficiency and return statistics: evidence from real estate and stock markets using a present value approach. *Real Estate Economics* 53, 171-195.
- Fuller, W. A. 1976. *Introduction to statistical time series*. New York. John Wiley and Sons.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457 (7232), 1012-14.

- Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37 (3), 424-438.
- Henderson, K. V. & Cowart, L. B. 2002. Bucking e-commerce trends: a content analysis comparing commercial real estate brokerage and residential real estate brokerage websites. *Journal of Corporate Real Estate* 4 (4), 375-385.
- Hohenstatt, R., Käsbauer, M., & Schäfers, W. 2011. "Geco" and its potential for real estate research: evidence from the U.S. housing markets. *Journal of Real Estate Research* 30 (4), 471-506.
- Hohenstatt, R., & Käsbauer, M. 2014. GECO's weather forecast for the UK housing market: to what extent can we rely on Google econometrics? *Journal of Real Estate Research* 36 (2), 253-281.
- Huoneistokeskus. 2011. Millainen on hyvä välittäjä? Haettu 11.2.2015 osoitteesta: <URL:http://yritys.huoneistokeskus.fi/kuukausitiedote.php?530>.
- Huoneistokeskus. 2012. Asuntomarkkinat kuukausikatsaus: Tammikuu 2012. Haettu 3.8.2015 osoitteesta: <URL:http://yritys.huoneistokeskus.fi/kuukausitiedote.php?16>.
- Huoneistokeskus. 2013. Asuntomarkkinat kuukausikatsaus: Toukokuu 2013. Haettu 3.8.2015 osoitteesta: <URL:http://yritys.huoneistokeskus.fi/kuukausitiedote.php?261>.
- International Telecommunications Union(ITU). World Telecommunication/ICT Indicators Database. Haettu 12.8.2015 osoitteesta: <URL:http://data.un.org/Data.aspx?q=haiti&d=ITU&f=ind1Code%3AI99H%3BcountryCode%3AHTI>.
- Johansen, S. 1988. Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12, 231-254.
- Johansen, S. & Juselius, K. 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52 (2), 169-210.
- Kasso, M. 2011. Ohjeita asuntokauppaan. Haettu 7.8.2015 osoitteesta: <URL:http://asunnot.oikotie.fi/ohjeita-asuntokauppaan>.
- Kouwenberg, R. & Zwinkels, R. 2014. Forecasting US housing market. *International Journal of Forecasting* 30, 415-425.
- Malik, T. M., Gumel, A., Thompson, H., L., Strome, T. & Salaheddin, M., M. 2011. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. *Canadian Journal of Public Health* 102 (4), 294-297.
- McLaren, N. & Shanbhogue, R. 2011. Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* nr. 2011 Q2, 134-140.
- Mellon, J. 2014. Internet search data and issue salience: the properties of Google Trends as a measure of issue salience. *Journal of Elections, Public Opinions and Parties* 24 (1), 45-72.
- Netmarketshare. Desktop Search Engine Market Share December, 2014. Haettu 2.10.2015 osoitteesta: <URL:https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0>.

- Oikarinen, E. 2005. Is housing overvalued in the Helsinki metropolitan area, *Keskusteluaiheita* no. 992, ETLA.
- Oikarinen, E. 2009. Household borrowing and metropolitan housing price dynamics—empirical evidence from Helsinki. *Journal of Housing Economics* 18, 126-139.
- Oikarinen, E. 2012. Empirical evidence on the reaction speeds of housing prices and sales to demand shocks. *Journal of Housing Economics* 21, 41-54.
- Perron, P. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361-1401.
- Proulx, R., Massicotte, P. & Pépino, M. 2014. Googling trends in conservation biology. *Conservation Biology* 28 (1), 44-51.
- Rochdi, K. & Dietzel, M. 2015. Outperforming the benchmark: online information demand and REIT market performance. *Journal of Property & Finance* 33 (2), 169-195.
- Schiller, R. J. 2007. Understanding recent trends in house prices and home ownership. NBER Working Paper No. 13553, Cambridge, MA: National Bureau of Economic Research.
- Scott, S. & Varian, H. 2014. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation* 5, 4-23
- Sims, C. 1980. Macroeconomics and reality. *Econometrica* 48, 1-48.
- Suhoy, T. 2009. Query indices and a 2008 downturn: Israeli data. Bank of Israel Discussion Paper 2009.06.
- StatCounter. GlobalStats: Top 5 Search Engines in Finland from Dec 2008 to Aug 2015. Haettu 4.2.2015 osoitteesta: <URL:http://gs.statcounter.com/#search_engine-eu-monthly-201001-201501>.
- StatCounter Global Stats, Top 5 Search Engines in Finland from Dec 2008 to Aug 2015. Haettu 12.8.2015 osoitteesta: <URL:http://gs.statcounter.com/#all-search_engine-FI-monthly-200812-201508>.
- Suomen Pankki. 2011. Finanssikriisin vaikutuksista Suomen talouteen. BoF Online 1/2011.
- Suomen Pankki. 2014. Talouden näkymät. Euro ja Talous 5/2014.
- Taylor, K. & McNabb, R. 2007. Business cycles and the role of confidence: evidence for Europe. *Oxford Bulletin of Economics and Statistics* 69 (2), 185-208.
- Theil, H. 1966. Applied economic forecasting. Chicago IL. The University of Chicago Press.
- Theil, H. 1971. Principles of econometrics. Amsterdam. North-Holland Publishing Co.
- Tilastokeskus. 2005. Väestön tieto- ja viestintätekniikan käyttö. Haettu 5.8.2015 osoitteesta:<URL :http://tilastokeskus.fi/til/sutivi/2005/sutivi_2005_2006-06-12_tie_001.html>.
- Tilastokeskus. 2011. Kotitalouksien varallisuustutkimus. Haettu 12.2.2015 osoitteesta: <URL:<http://www.stat.fi/til/vtutk/index.html>>.

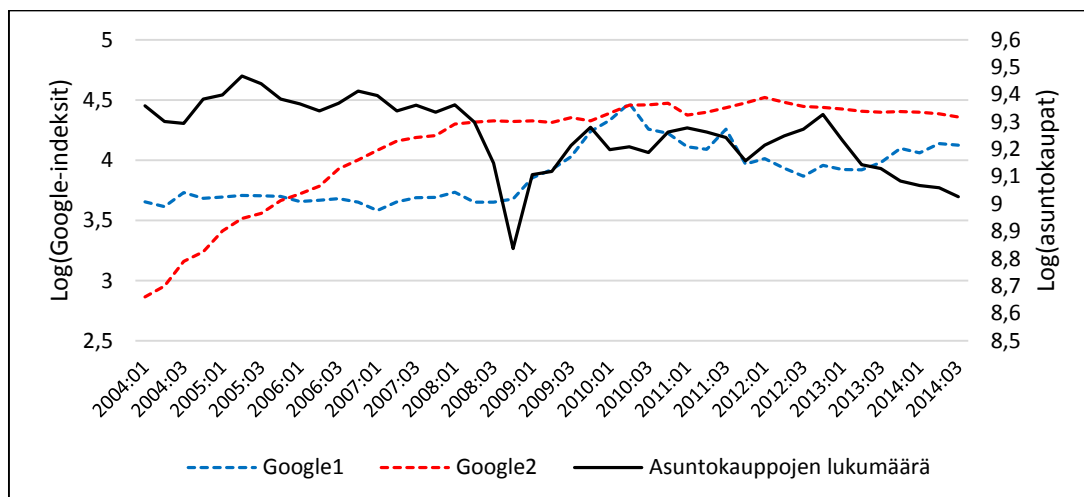
- Tilastokeskus. 2014a. Väestön tieto- ja viestintätekniiikan käyttö 2014. Haettu 10.10.2015 osoitteesta:
<URL:<http://www.stat.fi/til/sutivi/2014/index.html>>.
- Tilastokeskus. 2014b. Yleiskatsaus 2013, 1. Asuntokanta 2013. Haettu 1.11.2015 osoitteesta:<URL:http://tilastokeskus.fi/til/asas/2013/01/asas_2013_01_2014-10-16_kat_001_fi.html>.
- Tilastokeskus. 2015a. Kotitalouksien varallisuustutkimus. Haettu 10.5.2015 osoitteesta:<URL:http://www.stat.fi/til/vtutk/2013/vtutk_2013_2015-04-01_kat_002_fi.html>.
- Tilastokeskus. 2015b. Osakeasuntojen hinnat. Haettu 16.9.2015 osoitteesta:
<URL:http://www.stat.fi/til/ashi/2015/07/ashi_2015_07_2015-0828_laa_001_fi.html>.
- Tilastokeskus. 2015c. Tilastokeskus PX-Web-tietokannat: StatFin. Haettu 6.6.2015 osoitteesta: <URL:<http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/>>.
- Tilastokeskus. 2015d. Kerrostaloasuntojen kauppalukumäärät vuosina 2004-2007. Lähteenä Excel-taulukko, jonka toimitti Tilastokeskuksen aktuaari Petri Kettunen 11.2.2015.
- TNS Gallup. 2010. Asunnon osto- ja myyntiprosessitutkimus 2010. Haettu 11.2.2015 osoitteesta:
<URL:http://www.kvkl.fi/files/asunnon_osto_ja_myyntiprosessitutkimus_2010_kvkl%281%29.pdf>.
- TNS Gallup. 2013. Uutiskirje: Verkkopalveluiden rooli asuntoalalla suurempia kuin koskaan. Haettu 7.8.2015 osoitteesta: <URL:<https://www.tns-gallup.fi/uutiskirje/2013/verkkopalveluiden-rooli-asuntoalalla-suurempi-kuin-koskaan>>.
- TNS Gallup. 2014. Suomen web-sivutojen viikkoluvut 2014/1. Haettu 19.10.2015 osoitteesta: <URL:<http://tnsmatrix.tns-gallup.fi/public/>>.
- Tuhkuri, J. 2014. Big Data: Google-haut ennustavat työttömyyttä Suomessa. Etlan raportit no. 31.
- Vosen, S. & Schmidt, T. 2011. Forecasting private consumption: survey based indicators vs. Google Trends. *Journal of Forecasting* 30 (6), 565– 578.
- Vosen, S. & Schmidt, T. 2012. A monthly consumption indicator for Germany based on internet search query data. *Applied Economic Letters* 19, 683-687.
- Vänskä, L. 2013. "Kuplasta kuplaan" kerrostalohuoneistojen hintamuutokset Suomessa rahoitusmarkkinoiden vapauttamisesta nykypäivään. Jyväskylän yliopisto. Laskentatoimen pro gradu -työ.
- Wilcox, J. A. 2007. Forecasting components of consumption with components of consumer sentiment. *Business Economics* 42 (2), 36-46.
- Wu, L. & Brynjolfsson, E. 2013. The future of prediction: how Google searches foreshadow housing prices and sales. Julkaisematon.
- Yang, X., Pan, B., Evans, J. & Lv, B. 2015. Forecasting Chinese tourist volume with search engine data. *Tourism Management* 46, 386-397.
- Zivot, E. & Andrews, D. W. K. 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics* 10 (3), 251-70.

LIITTEET

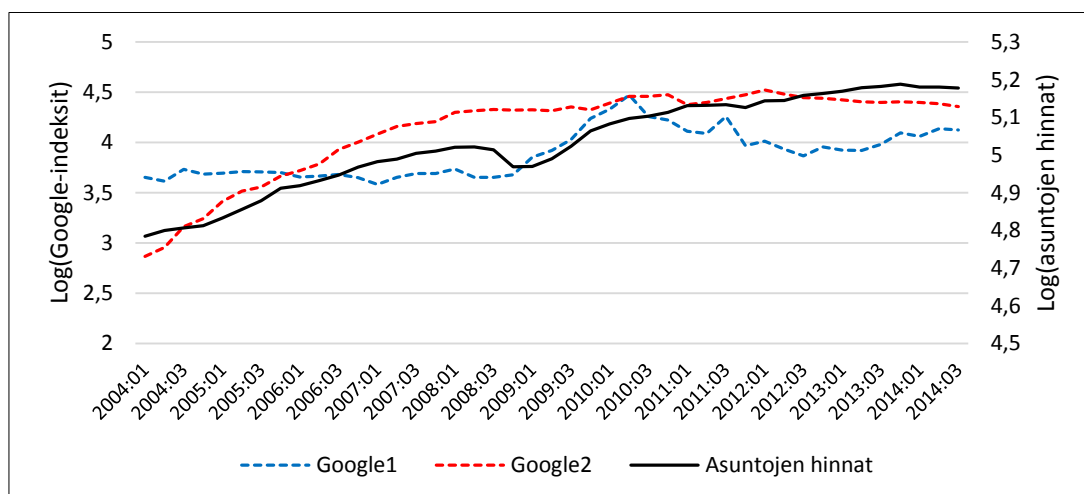
LIITE 1 Käytetyt Google-indeksit

Indeksi	Hakusanat
Google1-indeksi	asunnot + asunto + "myytävä asunto" + "myytävät asunnot" - vuokra - "vuokra-asunto" - vuokrattavat - "vuokra asunto" - "vuokra-asunnot"
Google2-indeksi	etuovi+oikotie+jokakoti

LIITE 2 Kausitasoitettujen asuntokauppojen lukumäärät sekä Google-indeksit



LIITE 3 Kausitasoitettujen asuntojen hinnat sekä Google-indeksit



LIITE 4 Zivot-Andrews-testin ehdottama rakenteellisen muutoksen ajankohta

Muuttujat	Viiveiden määrityskriteeri	
	BIC	MAXLAGS
Asuntokaupat (AK)	2008:02 (0)	2010:02 (2)
Asuntojen hinnat (AH)	2007:04 (1)	2008:02 (2)
Google1 (G1)	2009:01 (0)	2009:01 (2)
Google2 (G2)	2006:03 (0)	2006:03 (2)

LIITE 5 Differoitujen muuttujien tunnuslukuja

Muuttuja	Kes- kiarvo	Varianssi	Vinous	Huipuk- kuus	Ensimmäisen asteen autokorrelaatio
ΔAK (log)	-0,0079	0,007	-0,344	5,706	-0,07
ΔAH (log)	0,009	0,000	1,138	5,357	0,46***
$\Delta G1$ (log)	0,011	0,009	-0,596	1,829	-0,09
$\Delta G2$ (log)	0,036	0,004	0,681	0,997	0,50***
ΔY (log)	0,005	0,000	-0,615	1,452	0,13
ΔU (log)	0,001	0,001	1,657	3,673	0,25*
ΔI	-0,031	0,122	-3,521	17,38	0,46***

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 % (Ljung-Box).

Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

LIITE 6 Kontemporaalinen tarkastelu makrotalous huomioiden

A. Malli kaupoille (25) + makrotalous		B. Malli hinnoille (25) + makrotalous	
Selittävät muuttujat		Selittävät muuttujat	
$\log(\Delta AK)_{t-1}$	0,32* (0,18)	$\log(\Delta AH)_{t-1}$	0,58*** (0,18)
$\log(\Delta G1)$	0,06 (0,13)	$\log(\Delta G1)$	0,02 (0,02)
$\log(\Delta Y)$	1,27 (0,89)	$\log(\Delta Y)$	0,27* (0,14)
$\log(\Delta U)$	-0,55 (0,42)	$\log(\Delta U)$	0,11* (0,06)
ΔI	-0,21*** (0,06)	ΔI	-0,01 (0,01)
Vakiotermi	-0,02 (0,01)	Vakiotermi	0,00* (0,00)
Sopeutettu R ²	0,24	Sopeutettu R ²	0,31
AIC	-2,25	AIC	-5,93
BIC	-1,99	BIC	-5,68

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä estimaattien keskivirheet. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

LIITE 7 Granger-kausaisuustestien tulokset vuosidifferoiduille muuttujille

Paneeli A. Granger-kausaisuustestit asuntomarkkinamuuttujille ja Google1-indeksille

Selittävät muuttujat	Selitettävät muuttujat		
	log(Δ AK)	log(Δ AH)	log(Δ G1)
F-testin arvot			
log(Δ AK)	-	4,10*(1)	0,14 (1)
log(Δ AH)	1,22 (1)	-	0,52 (1)
log(Δ G1)	0,67 (1)	0,00 (1)	-
log(Δ AK)	-	0,94 (4)	0,70 (4)
log(Δ AH)	0,23 (4)	-	3,12**(4)
log(Δ G1)	0,81 (4)	1,05 (4)	-

Paneeli B. Granger-kausaisuustestit asuntomarkkinamuuttujille ja Google2-indeksille

Selittävät muuttujat	Selitettävät muuttujat		
	log(Δ AK)	log(Δ AH)	log(Δ G2)
F-testin arvot			
log(Δ AK)	-	4,10*(1)	4,70***(4)
log(Δ AH)	1,22 (1)	-	0,73 (1)
log(Δ G2)	0,23 (4)	0,26 (1)	-
log(Δ AK)	-	0,94 (4)	4,70***(4)
log(Δ AH)	0,23 (4)	-	1,73 (4)
log(Δ G2)	0,23 (4)	0,12 (4)	-

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä VAR-malleissa käytetyt viivepituudet, jotka määritettiin Bayesilaisella informaatiokriteerillä (ylemmät tulokset) tai tekemällä alkuoletus mahdollisesta viivepituudesta (alemmat tulokset). Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

LIITE 8 Kontemporaalinen tarkastelu vuosidifferoiduilla muuttujilla

Paneeli A. Asuntokaupat			
Malli	(24)	(25)	(25) + makrotalous
Selittävät muuttujat			
log(Δ AK)t-1	-0,02 (0,17)	0,03 (0,16)	0,34* (0,18)
log(Δ G1)		0,33** (0,14)	0,15 (0,14)
log(Δ Y)			0,87 (0,96)
log(Δ U)			-0,49 (0,45)
Δ I			-0,21*** (0,06)
Vakiotermi	-0,01 (0,02)	-0,01 (0,02)	-0,00 (0,02)
Sopeutettu R ²	-0,03	0,08	0,30
AIC	-1,05	-1,13	-1,33
BIC	-0,96	-1,00	-1,01
Paneeli B. Asuntojen hinnat			
Malli	(24)	(25)	(25) + makrotalous
Selittävät muuttujat			
log(Δ AH)t-1	0,50*** (0,15)	0,48*** (0,14)	0,67*** (0,18)
log(Δ G1)		0,04* (0,02)	0,03 (0,02)
log(Δ Y)			0,19 (0,14)
log(Δ U)			0,14** (0,05)
Δ I			-0,01 (0,01)
Vakiotermi	-0,00 (0,00)	-0,00* (0,00)	-0,00 (0,00)
Sopeutettu R ²	0,25	0,30	0,41
AIC	-4,98	-5,06	-5,15
BIC	-4,89	-4,92	-4,89

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä estimaattien keskivirheet. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.

LIITE 9 Granger-kausaisuustestien tulokset X11 Census-menetelmällä kausitasoitetuille muuttujille aikavälillä 1:2004–1:2008

Paneeli A. Granger-kausaisuustestit asuntomarkkinamuuttujille ja Google1-indeksille

Selitettävät muuttujat			
Selittävät muuttujat	log(Δ AK)	log(Δ AH)	log(Δ G1)
F-testin arvot			
log(Δ AK)	-	1,93 (1)	0,50 (4)
log(Δ AH)	2,03 (1)	-	2,63 (1)
log(Δ G1)	7,46*(1)	0,02 (1)	-

Paneeli B. Granger-kausaisuustestit asuntomarkkinamuuttujille ja Google2-indeksille

Selitettävät muuttujat			
Selittävät muuttujat	log(Δ AK)	log(Δ AH)	log(Δ G2)
F-testin arvot			
log(Δ AK)	-	1,93 (1)	0,00 (1)
log(Δ AH)	2,03 (1)	-	0,33 (1)
log(Δ G2)	9,42***(1)	0,00 (1)	-

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä VAR-malleissa käytetyt viivepituudet. Muuttujien lyhenteiden selitykset on esitetty taulukossa 4.

LIITE 10 Kontemporaaლის Google1-indeksin kyky selittää muutoksia reaalisissa asuntojen hinnoissa

Malli	(24)	(25)	(25) + makrotalous
Selittävät muuttujat			
log(Δ AH) _{t-1}	0,52*** (0,14)	0,50*** (0,13)	0,56*** (0,14)
log(Δ G1)		0,04** (0,02)	0,03* (0,02)
Vakiotermi	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Kontemporaaლის makrotaloudelliset muuttujat			x
Sopeutettu R ²	0,25	0,31	0,42
AIC	-5,81	-5,87	-5,97
BIC	-5,73	-5,75	-5,72

*, ** ja *** kuvaavat tilastollista merkitsevyyttä tasoilla 10 %, 5 % ja 1 %. Suluissa ovat esitettyinä estimaattien keskivirheet. Muuttujien lyhenteiden selitykset ovat esitettyinä taulukossa 4.