

Kansantaloudellinen aikakauskirja – 115. vsk. – 1/2019

Sosiaalisen median ennustekyky kaupan myynnissä – esimerkkinä veganismi ja vegaanisit ruuat

Veikko Isotalo, Salla-Maaria Laaksonen, Essi Pöyry ja Piia Jallinoja

Sosiaalisen median aineistoja on pyritty hyödyntämään ennustemalleissa esimerkiksi vaalitulosten, työttömyysasteen, osakekurssien tai kuluttajatuotteiden kysynnän ennakoinnissa. Laajempia kulutusilmiöitä on kuitenkin harvemmin pyritty ennustamaan sosiaalisen median avulla. Tämä tutkimus keskittyy veganismiin, viime vuosina pinnalla olleeseen kulutusilmiöön, vertaamalla sitä koskevien sosiaalisen median viestien määrää vegaanisten kasvimaitotuotteiden myyntiin. Analysoimalla viikoittaista myyntimäärää ja sosiaalisen median bakusanaryppäiden osuimien määrää vuodesta 2012 vuoteen 2016 autoregressiivisen vektorivirheenkorjausmallin (VECM) avulla tämä tutkimus osoittaa, että kasvimaitotuotteiden myynnillä sekä sosiaalisen median viesteillä on olemassa pitkän aikavälin riippuvuusuhde. Yleistä ilmiötä, veganismia, koskevat viestit selittivät kasvimaitotuotteiden myyntiä, mutta kyseisiä tuotteita koskevat viestit eivät. Myynnin ennustamisessa toimii kuitenkin parhaiten pelkkään myyntiaineistoon pohjautuva SARIMAX-ennuste.

Sosiaalisen median käytön lisääntynyt suosio, alustojen moninaisuus ja sosiaalisen median sisällön räjähdysmäinen kasvu on herättänyt niin tutkijoiden, liikkeenjohdon kuin päättäjien mielenkiinnon sen mahdolliseen hyödyntämiseen erilaisten kuluttaja- ja kansalaisteemojen ennustamiseen. Sosiaalista mediaa on käytetty

selittämään tai ennustamaan esimerkiksi vaalituloksia (DiGrazia ym. 2013), työttömyysastetta (Tuhkuri 2016), influenssan leviämistä (Aramaki ym. 2011; Culotta 2010) ja erilaisten tuotteiden kysyntää (Dewan ja Ramaprasad 2014; Stephen ja Galak 2012). Yleinen konsensus on, että monissa teemoissa sosiaalinen media pa-

VTM Veikko Isotalo (veikko.isotalo@helsinki.fi) on väitöskirjatutkija Helsingin yliopiston valtiotieteellisessä tiedekunnassa (valtio-oppi) ja tutkimusharjoittelija Kuluttajatutkimuskeskuksessa. VTT Salla-Maaria Laaksonen (salla.laaksonen@helsinki.fi) ja KTT Essi Pöyry (essi.poyry@helsinki.fi) ovat tutkijatohtoreita Helsingin yliopiston Kuluttajatutkimuskeskuksessa. Professori Piia Jallinoja (piia.jallinoja@tuni.fi) on terveys sosiologian professori Tampereen yliopistossa. Tämä tutkimus on toteutettu osana Tekes/Business Finlandin rahoittamaa Smarter Social Media Analytics -tutkimushanketta (4161/31/2016). Kiitämme kollegoitamme Juho Pääkköstä, Arto Kekkosta, Mikko Jauhoa ja Antti Salovaaraa avusta aineiston esikäsitelyssä ja tutkimuksen ideoinnissa. Kiitämme myös anonyymia lausunnonantajaa perusteellisista ja rakentavista kommentista. Kiitämme aineistosta ja yhteistyöstä Futusome Oy:tä ja S-ryhmää.

rantaa erilaisten ennustemallien tarkkuutta, mutta on harvoin yksinään riittävä mittari luotettavalle ennustamiselle (Dellarocas ym. 2007; Dhar ja Chang 2009; Gayo-Avello 2013). Tapaesimerkkejä sosiaalisen median keskustelujen vaikutuksesta kuitenkin on; esimerkiksi Mutti-tomaattimurska loppui kauppojen hyllyiltä tammikuussa 2014 julkkiskokin twiitin seurauksena (ks. Pöyry ym. 2018).

Kuluttajatuotteiden myynnin ennustamisessa sosiaalisen median data on osoitettu toimivaksi erityisesti tietyissä tuoteryhmissä. Esimerkiksi elokuvalippujen menekkiä koskevat ennustemallit usein paranevat, kun otetaan huomioon sosiaaliseen mediaan tuotettujen elokuva-arvioiden määrä, tahti ja sävy (Asur ja Huberman 2010; Chintagunta ym. 2010; Moon ym. 2010; Onishi ja Manchanda 2012). Lisäksi tiettyjen brändien myyntiä tai niitä tuottavien yritysten osakekurssien ennustemalleja on parannettu ottamalla huomioon sosiaalisen median keskustelujen määrä tai sävy (Bollen ym. 2011; Yu ym. 2013; Luo ym. 2013; Stephen ja Galak 2012). Tyypillisesti näissä tutkimuksissa on keskitytty yhteen sosiaalisen median palveluun, useimmiten Twitteriin, ja haettu sosiaalisen median sisältöjä tunnistettavilla brändinimillä.

Kauppan alan kontekstissa on kuitenkin huomattava, että useimmat yritykset haluaisivat pystyä ennakoimaan laajempia kuluttajatrendejä kuin yksittäisten tuotteiden myyntiä. Näin on esimerkiksi ruokakaupassa, jossa eri brändien ja tuotetyyppien välillä saattaa olla suuria eroja niiden tunnettuudessa ja kuluttajien huomiosta – monista arkisista tuotteista ei keskustella aktiivisesti sosiaalisessa mediassa, vaikka niitä ostetaan tiuhaan. Aiempi tutkimus ei ole kuitenkaan selvittänyt, voisiko sosiaalista mediaa hyödyntää ennustamaan kokonaisten

tuotekategorioiden myyntiä niitä koskevien kategoriataason viestien avulla ottamatta huomioon kategoriaan kuuluvien yksittäisten brändien nimiä. Haluamme tutkia aihetta pelaamalla sekä kategoriaan liittyvän ”ilmiökeskustelun” että sen tuotteita koskevien keskustelujen määrää kategorian tuotteiden myynnin määrään. Lähestymme tutkimusongelmaa täsmällisemmin kahden tutkimuskysymyksen kautta: 1) Minkälaisia tilastollisesti merkitseviä dynaamisia yhteyksiä sosiaalisen median keskusteluja ja tuotekategorian myyntiä kuvaavien aikasarjojen väliltä löytyy? 2) Miten ilmiö- ja tuotekeskustelujen määrä sosiaalisessa mediasa ennakoii kategorian tuotteiden myynnin määrää?

Aihetta tutkitaan käyttämällä tutkimustapauksena veganismia, jota voidaan luonnehtia nousevaksi kulutusilmiöksi (Jallinoja ym. 2019a). Veganismi – eläinkunnan tuotteiden välttäminen ruokavaliossa – sopii hyvin tutkitavaksi ilmiöksi, koska sitä käsitteleviä sosiaalisen median viestejä on runsaasti, se on yksiselitteisesti tunnistettavissa selkeillä hakusanoilla ja kyseiseen ruokavalioon voidaan liittää useita tuotteita, jotka eivät ole perinteinen osa suomalaista keittiötä tai muita yleisimpiä ruokavalioita. Tutkimusaineistona käytetään viikoittaisten avointen suomenkielisten sosiaalisen median viestien määrää sekä suomalaisen, valtakunnallisen elintarvikemyymäläketjun myymälöissä myytyjen kasvimaitotuotteiden myyntimäärää vuoden 2012 alusta vuoden 2016 loppuun. Ilmiökeskusteluja etsitään veganismiin liittyvillä hakusanoilla ja tuotekeskusteluja kasvimaitotuotteisiin liittyvillä hakusanoilla. Analyysimenetelmänä käytetään vektoriautoregressiota (VAR) ja sen laajennusta VEC-mallinnusta, joka huomioi VAR:sta poiketen myös muuttujien pitkän aikavälin suhteet.

1. Kirjallisuuskatsaus

1.1 Sosiaalisen median hyödyntäminen ennustamisessa

Sosiaalinen media on kattotermi, joka viittaa joukkoon erilaisia käyttäjäsisällön luomisen ympärille rakentuneita verkkopalveluita (Laaksonen ym. 2013). Julkisessa keskustelussa termi viittaa tyypillisesti Facebookin ja Twitterin kaltaisiin sosiaalisiin verkostopalveluihin, mutta myös sosiaalisen median vanhempia muotoja kuten blogeja ja keskustelupalstoja käytetään yhä runsaasti (SVT 2018). Tutkimusaineistona sosiaalinen media näyttää ennen kaikkea kiinnostavana, kuluttajien luonnollisessa arjen kontekstissa tuottamana datamassana, jota voidaan hyödyntää pitkittäisinä, aikaan kiinnitettyinä aineistoina. Tällaisten laajojen tekstiaineistojen saatavuus on viimeisen vuosikymmenen ajan herättänyt kiinnostusta niiden hyödyntämiseen niin tutkimuksessa kuin käytännön sovellutuksissakin (Nelimarkka ja Laaksonen 2018).

Tutkimuksissa on osoitettu esimerkiksi, että elokuvan menestymistä lippuluukulla voidaan ennakoita elokuvaa koskevien twiittien ilmestymistiheydellä (Asur ja Huberman 2010), ja että Twitterin viestien sentimentin avulla voi selittää pörssikurssien vaihteluja (Bollen ym. 2011; Yang ym. 2015). Toisaalta perinteisten markkinamittareiden merkitys ei näytä olevan heikentynyt – esimerkiksi mainonnan määrä ja maininnat uutismediassa ovat usein parantaneet ennustemalleja (Dewan ja Ramaprasad 2014; Dhar ja Chang 2009; Stephen ja Galak 2012; Onishi ja Manchanda 2012). Poliitiikan tutkimuksen parissa on tehty lukuisia analyyseja sosiaalisen median viestien kyvystä ennakoita vaalituloksia. Johtopäätökset ovat varsin

ristiriitaisia. Jotkut analyytit osoittavat sosiaalisella medialla olevan ennustevoimaa (Tumasjan 2010; DiGrazia ym. 2013), mutta löydöksiä on myös kritisoitu (Jungherr ym. 2011; Murthy 2015; Gayo-Avello 2013).

Sosiaalisen median viestien määrän lisäksi myös muunlaisia verkkokäyttäytymisen mittareita on hyödynnetty eri asioiden ennustamisessa tai suosion selittämisessä. Esimerkiksi Yaseri ja Bright (2016) osoittivat, että poliittisten aiheiden Wikipedia-sivujen kävijämäärät toimivat indikaattorina äänestäjien kiinnostuksesta ja siksi niiden avulla voi ennakoita vaalimenestystä. Vastaavasti O'Connor (2013) havaitsi, että brändien Facebook-sivujen tykkääjien määrä toimii indikaattorina brändin omistavan yrityksen osakkeiden arvosta. Google-hakuja hyödyntävistä tutkimuksista tunnetuin esimerkki lieenee Google Flu Trends (Ginsberg ym. 2009), joka ennakoiti influenssaepidemian puhkeamista Googlessa tehtyjen hakujen perusteella. Suomessa Gencoglu ja Ermes (2018) ovat toistaneet flunssaennustemallin vertaamalla Instagram-kuvien sisältöjä Terveiden ja hyvinvoinnin laitoksen ennusteisiin. Tuhkuri (2014, 2016) taas paransi Elinkeinoelämän tutkimuslaitoksen työttömyysastetta koskevia mallinnuksia liittämällä niihin indikaattoreita työttömyyttä koskevien Google-hakujen määristä.

Edellä mainittujen elokuva- ja osakekurssitutkimusten lisäksi pitkittäisten sosiaalisen median keskusteluaineistojen käyttö kuluttajaliiketoiminnan ennakoinnissa on ollut varsin rajoittunutta. Olemassa oleva tutkimus on keskittynyt lähinnä brändejä koskevien sosiaalisen median viestien vaikutukseen niiden myyntiin tai muihin vastaaviin mittareihin. Esimerkiksi Stephen ja Galak (2012) tutkivat erästä mikro-lainapalvelua koskevien blogi- ja foorumiviestien määrän vaikutusta uusiin lainahakemuk-

siin 14 kuukauden aikana ja totesivat, että sosiaalisen median viesteillä oli hakemusten määrään positiivinen vaikutus. Vastaavasti Kim ja Hanssens (2017) osoittivat elokuvia ja videopelejä koskevien blogikirjoitusten vaikuttavan kuluttajien kiinnostukseen pitkäkestoisemmin kuin mainostaminen. Dewan ja Ramaprasad (2014) sen sijaan havaitsivat, että kahden vuoden aikana viikoittaisten blogiviestien määrä tietystä musiikkialbumeista tai -kappaleista ei ennakoanut näiden viikoittaista myyntiä. Tutkijat spekulivat, että blogiviestit olivat lisänneet ilmaista musiikin kuuntelua netissä, ja siksi vaikutusta myyntiin ei ollut.

Näissä tutkimuksissa on tutkittu yksittäisten, selkeästi tunnistettavien brändien myyntiä sekä näitä brändejä koskevia sosiaalisen median viestejä. Tietääksemme aiempi tutkimus ei kuitenkaan ole ottanut huomioon, että joihinkin tuotteisiin liittyvä ylätasoinen kategoria (esim. kasvimaidot) tai ilmiö (esim. veganismi) voi olla kuluttajille relevantti keskustelun taso (ei siis ainoastaan brändit), ja niitä koskevien sosiaalisen median viestien lisääntyminen saattaa myös vaikuttaa tuotteiden myyntiin. Lisäksi elokuvia ja muita vastaavia viihdetuotteita koskevan tutkimuksen ollessa erittäin runsasta on tärkeää, että tutkimuksessa otetaan huomioon myös muita tuoteryhmiä, joita koskeva sosiaalisen median kirjoittelu voi myös olla erilaista.

1.2 *Veganismi 2010-luvulla*

Tämän tutkimuksen tavoitteena on analysoida sosiaalisen median keskusteluteemojen yleisyyden ja niitä koskevien kuluttajatuotteiden myyntimäärien yhteyttä. Valitsimme tutkimuskohteeksi veganismin ja vegaaniset kasvimaitotuotteet.

Vegaaniruokavalioon ei kuulu mitään eläinkunnan tuotteita, kuten lihaa, kalaa, maitotuotteita tai munia. Tyypillisesti vegaanit välttävät eläinkunnan tuotteita kaikilla elämäntilanteilla, kuten vaatteissa, kosmetiikassa ja lääkkeissä (Meindersma 2008) ja painottavat tasavertaista suhdetta ihmisten ja muiden eläinten välillä (Francione 2010). Veganismissa motivoivat tyypillisimmin eläinten kaltoinkohtelu tuotantotiloilla, ja seuraavaksi yleisimmän huoli ilmaston tilasta ja tämän jälkeen terveyskysymykset (Kerschke-Risch 2015). Euroopassa ja Pohjois-Amerikassa vegaanien osuus väestöstä on pysynyt varsin alhaisena. Sekä Suomessa että Yhdistyneessä kuningaskunnassa vuonna 2016 runsas prosentti (Jallinoja, ym. 2019b; The Vegan Society 2018) ja Ruotsissa vuonna 2017 kolme prosenttia ilmoitti olevansa vegaaneja (Djurens Rätt 2017).

2010-luvulla useiden mittareiden mukaan kiinnostus vegaanisia tuotteita ja vegaanista ruokavalioita kohtaan on kasvanut (Jallinoja ym. 2019b). Vegaanisten maitoa korvaavien juomien myynti kaksinkertaistui vuosien 2009 ja 2015 välillä (Whipp 2016). Suomessa Kesko raportoi näiden tuotteiden myynnin kasvaneen 47 prosenttia yhden vuoden aikana ja falafelituotteiden myynnin kasvaneen 305 prosenttia (Kesko 2017). Ruotsissa kasvissyönnistä kiinnostuneiden ei-kasvissyöjien osuus kasvoi 2009–2017 26 prosentista 47 prosenttiin (Djurens Rätt 2015, 2017). Kuvaavia ovat myös elintarviketeollisuuden vegaaniset innovaatiot – Nyhtökaura, Härkis, McVegan – jotka ovat saaneet runsaasti mediahuomiota. Yhdistyneessä kuningaskunnassa vuosien 2012 ja 2016 välillä uusien vegaanisten tuotteiden määrä kasvoi 185 prosenttia (Hancox 2018).

Siinä missä aikaisemmin veganismi on yhdistetty askeettisuuteen ja vegaanista ruokaa

on pidetty tylsänä ja mauttomana (Veron 2016), 2010-luvulla mielikuvat veganismista ovat myönteisempiä (Jallinoja ym. 2019a). Veganismin rinnalle on tullut lukuisia joustavuutta painottavia käsitteitä ja ruokavalioita, kuten *eating vegan before six o'clock* (Bittman 2013), *meatless Monday* (Morris 2018), ”fleksaaminen” ja erilaiset kasvissyöntihaasteet. Esimerkiksi Oikeutta eläimille -yhdistyksen kehittämä Vegaanihaaste, eli haaste noudattaa vegaanista ruokavaliota tammikuussa, on ollut viime vuosina suosittu Suomessa (Kallionpää 2016). Näin ollen on mahdollista, että koko vegaanituotteiden kategoria olisi hyötynyt sitä koskevasta myönteisestä huomiosta, esimerkiksi sosiaalisessa mediassa.

Sosiaalista mediaa on hyödynnetty jonkin verran ihmisten ruokailutottumuksia ja ruokaan liittyviä käsityksiä koskevissa tutkimuksissa. Instagram-kuvia tutkittaessa on muun muassa saatu selville, että kohtuullisen terveellisiä ruokia esittävät kuvat ovat tykkäyksillä ja kommenteilla mitattuna kaikkein suosituimpia (Sharma ja De Choudhury 2015). Terveellisiä ruokia koskevat viestit ovat myös yleisempiä hyväosaisilla alueilla kuin muilla alueilla (Abbar ym. 2015; De Choudhury ym. 2016). Näiden tutkimusten mukaan sosiaalisen median dataa voi hyödyntää parantamaan alueellisten lihavuus- ja diabetesennusteiden tarkkuutta ja paikallistamaan niin sanottuja ruoka-aavikkoja, eli alueita, joissa on heikko pääsy terveelliseen ja kohtuuhintaiseen ruokaan. Lisäksi Fried ym. (2014) väittävät, että Twitteristä pystyy löytämään tiettyjä kielellisiä johdonmukaisuuksia alueellisista ruokailutottumuksista ja terveyskäyttäytymisestä.

Vaihtelevista tutkimustuloksista huolimatta sosiaalisen median tutkiminen on tärkeää, koska sillä on nykyään erittäin suuri rooli kulutta-

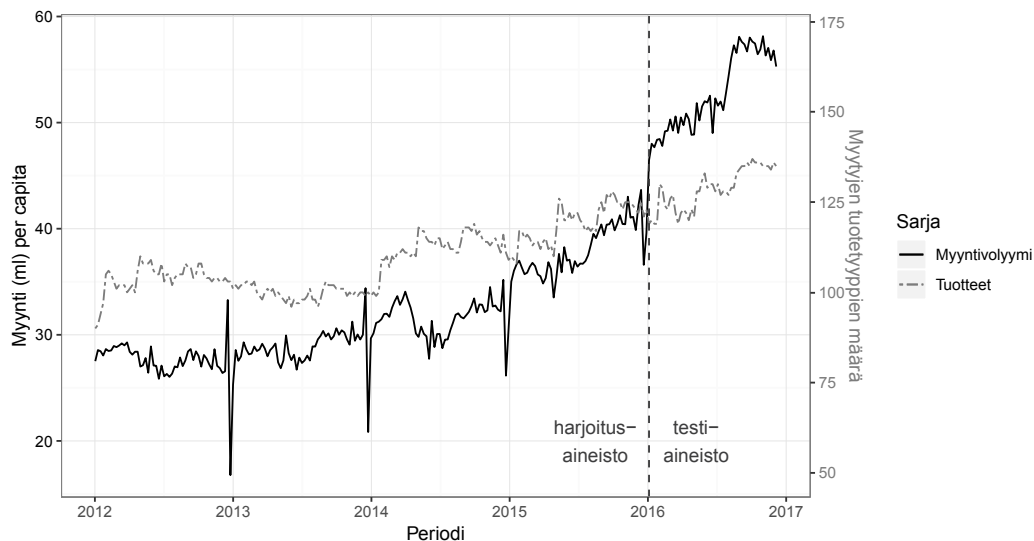
jen ja kansalaisten tiedon ja vaikutteiden saamisessa. Esimerkiksi rokotuskriittisyyden on osoitettu leviävän nimenomaan sosiaalisen median kanavissa (Jones ym. 2012; Wilson ja Keelan 2013), joten niiden huomioiminen kuluttajailmiöitä tutkittaessa on olennaista. Vegaanisen ruokavalion suosion leviämistä ei ole tiettävästi aiemmin tutkittu sosiaalisen median avulla.

2. Menetelmä

2.1 Aineisto

Aineisto kerättiin vuosilta 2012–2016 viikkotasolla (pl. vuoden 2016 viimeiset kolme viikkoa, yhteensä 258 viikkoa). Myyntidatan osalta käytämme S-ryhmän myymälöiden (Prisma, S-Market, Alepa, Sale) myyntitietoja valittujen kasvimaitotuotteiden osalta koko Suomessa. S-ryhmän markkinaosuus Suomessa vuosina 2012–2016 oli 45,6–47,2 % (Päivittäistavara-kauppa ry 2013, 2016), joten aineisto kertoo melko luotettavasti tutkittavien tuotteiden suosiosta koko Suomen väestön keskuudessa. Aineistoon kuului 245 eri tuotekoodilla olevaa tuotetta 28 eri tuottajalta tai maahantuojalta, ja se sisälsi kasvimaitoja, ruoanlaittoon tarkoitettuja kasvipohjaisia valmisteita (esim. kerman kaltaisia tuotteita), kasvipohjaisia jogurtteja, rahkoja, ynnä muita välipalatuotteita, kasvipohjaisia juustoja ja levitteitä, sekä kasvipohjaisia jäätelöitä. Aineisto perustui toisaalta tunnettujen tuottajien ja brändien listaan ja toisaalta relevanttien tuotteiden etsintään koko maitokategoriasta. Aineiston arvioidaan kattavan haluttu tuotekategoria hyvin, mutta joitakin yksittäisiä, pienempien tuottajien tuotteita puuttuneen. Kaikki aineiston tuotteet olivat ve-

Kuvio 1. Kasvimaitotuotteiden myyntimäärä viikoittain 2012–2016



Lähde: S-ryhmä/SOK-yhtymä.

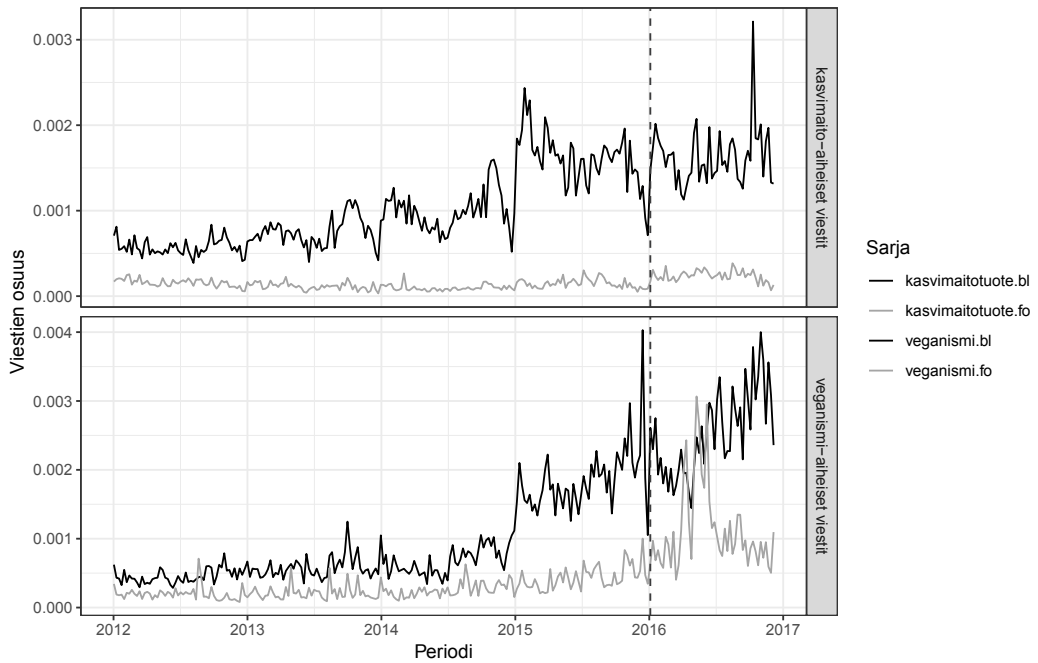
gaanisista, eli ne eivät sisältäneet eläinkunnan raaka-aineita, kuten lehmän maitoa. Kuvio 1 esittelee aineiston tuotteiden kokonaismyymintämäärät millilitroissa tai grammoissa väestöön suhteutettuna sekä myytyjen tuotenimikkeiden kappalemäärät viikoittain 2012–2016. Kasvimaitojen myyntivolyyymissa on tapahtunut 109 prosentin kasvu kun verrataan satunnaista loppusyksyn viikkoa (vko 49) vuosina 2012 ja 2016. Myytyjen tuotenimikkeiden myynti kasvoi samana aikana 31 prosenttia.

Sosiaalisen median aineisto koottiin hyödyntäen suomalaisen sosiaalisen median sisältöjä keräävän yrityksen Futusome oy:n rajapintaa. Futusome kerää kattavasti suomalaisia sosiaalisen median viestejä tietokantaan, johon kertyy päivittäin noin 300 000–500 000 uutta

viestiä. Palvelu kerää avoimia sosiaalisen median viestejä Facebookista, keskustelupalstoilta eli foorumeista, Twitteristä, Instagramista, uutiskommenteista, blogeista, YouTubeista ja Pinterestistä (Laaksonen ja Pöyry 2018).

Haku tehtiin käyttämällä sanalistoja, jotka koottiin tutkijoiden yhteistyössä erityisesti ruokasosiologiaan erikoistuneen kirjoittajan asiantuntemuksen avulla. Omat hakunsa tehtiin kummallekin aiheelle: ilmiötason viesteille (veganismi) sekä tuotetason viesteille (kasvimaitotuotteet). Ilmiötason viesteihin luettiin kuuluvan sosiaalisessa mediassa 2012–2016 julkaistut julkiset viestit, jotka sisälsivät sanan ”vegaani”, ”vegaaninen”, ”vegaaniruoka” tai ”veganismi”. Kasvimaitotuotteita koskevia viestejä etsittiin yhteensä 30 hakusanalla kuten

Kuvio 2. Kasvimaito- ja veganismiaiheisten viestien normalisoidut määrät blogeissa ja foorumeissa viikoittain vuosina 2012–2016



Lähde: Futosome Oy.

”kauramaito”, ”soijamaito”, ”kookosmaito” ja ”vegaanijuusto”.¹ Jälkimmäinen hakusanalista perustui niihin tuotenimikkeisiin, jotka sisältyivät myyntidataan. Viestit, jotka sisälsivät useamman kuin yhden samaan aiheeseen kuuluvan hakusanan, laskettiin aineistoon vain kerran. Jos viesti sisälsi molempiin aiheisiin

¹ Kaikki hakusanat aakkosjärjestyksessä: havregurt, juustonkorvike, kasvimaito, kaurafraiche, kaurajugurtti, kaurajuoma, kaurajäätelö, kauramaito, kauravalmiste, kookosjuoma, kookoskerma, kookosmaito, maidonkorvike, mantelijuoma, mantelikerma, mantelimaito, pähkinäjuoma, pähkinämaito, riisijuoma, riisimaito, soijajugurtti, soijajuoma, soijajäätelö, soijakerma, soijamaito, soijavalmiste, soijavanukas, soygurt, vegaanijuusto, vegaanijäätelö

kuuluvia hakusanoja, se laskettiin kuuluvan niihin molempiin.

Saadaksemme mahdollisimman tasalaatuisia aineistoa rajasimme sen avoimen sosiaalisen median viesteihin eli blogi- ja foorumiviesteihin. Rajasimme ulos sosiaaliset verkostoitumispalvelut Facebookin, Twitterin ja Instagramin, sillä Futosome on kerännyt niitä kattavasti vasta vuodesta 2014 alkaen, eikä datan kerääminen niistä jälkikäteen ole mahdollista (Mayr ja Weller 2017). Muut alustat jätettiin tarkastelun ulkopuolelle, koska niiden viestimäärät olivat merkittävästi vähäisemmät. Hakumme tuotti yhteensä 50 886 kasvimaitoaiheista ja 88 332 veganismiaiheista viestiä. Kuvion 2 ylä-

paneeli esittää kasvimaitoaiheisten ja alapaneeli veganismiaiheisten viikoittaiset määrät vuosina 2012–2016 blogeissa ja foorumeissa. Jotta mahdolliset muutokset tietokannan keruussa eivät vaikuttaisi aineistoon, kaikki muuttujat normalisoitiin suhteessa tietokannan päivittäisiin alustakohtaisiin kokonaisviestimääriin.

2.2 Analyysi

Käytimme analyysimenetelmänä vektoriautoregressiota ja sen sovellusta vektorivirheenkorjausmallia (Brandt ja Williams 2007). Autoregressiomenetelmä perustuu ajatukseen, jonka mukaan muuttujan ajanhetken t - i arvoja, jossa $i = 1, 2, \dots, p$, voidaan käyttää ennustamaan saman muuttujan ajanhetken t arvoa. Esimerkiksi kasvimaitojen myynti annettuna viikkona otetaan huomioon ennustettaessa seuraavan ja sitä seuraavien viikkojen myyntiä. Vektoriautoregressiossa (VAR) mallinnetaan useiden muuttujien ajanhetken t arvoja kaikkien mallin endogeenisten muuttujien t - i arvojen perusteella (Pecican 2010; Lütkepohl 2009). Menetelmä sopii aikasarjamallinnukseen, kun muuttujien välisille suhteille ei voida tai haluta määrittää rakenteita *ex ante* esimerkiksi tarkastelemalla jotain muuttujaa ainoastaan riippuvana muuttujana (Dewan ja Ramaprasad 2014; Brandt ja Williams 2007). Kasvimituotteiden myynnin ja sosiaalisen median viestien määrän kohdalla esimerkiksi hyväksytään, että viestien määrän muutos voi vaikuttaa myyntiin, mutta myös myynnin muutos voi vaikuttaa viestien määrään. VAR-mallien muuttujia käsitellään siis endogeenisinä muuttujina.

Jos tutkittavat muuttujat ovat vahvasti yhteisintegroituvia, käytetään VAR:n laajennusta vektorivirheenkorjausmallia (Vector Error

Correction Model VECM, ks. Lütkepohl 2004). VEC-mallinnuksessa estimoidaan aineistolle ensin yhteinen stokastinen trendi kuvaamaan muuttujien pitkän aikavälin suhdetta, minkä jälkeen lyhyen aikavälin joustoja estimoidaan regressiomallien ja virheenkorjaustermien avulla (Brandt ja Williams 2007). VEC on siis tehokas menetelmä muuttujien välisten suhteiden ja kausaliteettisuunnan tutkimiseen. VAR- ja VECM-mallinnuksiin on mahdollista lisätä myös eksogeenisiä kontrollimuuttujia, jolloin käytetään toisinaan nimitystä VARX tai VECMX.

Tässä tutkimuksessa muodostimme VARX-malleja käyttämällä taulukossa 1 listattuja aikasarjamuuttujia sekä eksogeenisiä kontrollimuuttujia ja sovitimme mallit ensin harjoitusaineistoon. Mahdollisina kontrollimuuttujina kokeiltiin myytyjen tuotenimikkeiden määrää, pääsiäistä, joulua sekä Vegaanihaaste-kautta (tammi- ja joulukuun 2015 ja 2016). Joulut mallinnettiin erikseen vuosille 2012–2013 sekä 2014–2015, sillä myyntikäyrän visuaalisen tarkastelun perusteella niiden vaikutus tuotteiden myyntiin oli selkeästi erilainen. Havainto sai myös tukea aikasarjamallien korjattujen selitysasteiden noustua kahden joulukontrollin malleissa. Joulujen osalta huomioitiin binäärimuuttujilla jouluviikon lisäksi joulua edeltävä ja joulua seuraava viikko, jolloin kuluttajien ostokäyttäytyminen muuttuu. Pääsiäisviikolla tai tuotteiden määrällä ei ollut vastaavaa merkitsevää vaikutusta yhdessäkään mallissa, joten pudotimme ne pois. Samoin kasvimituotteita koskevat foorumikeskustelut sekä veganismia koskevat blogiviestit rajautuivat pois valitusta VAR-mallista. Analyysin vaiheet on esitetty taulukossa 2.

Mallinnuksessa käytettiin R-pakettia *vars* (Pfaff 2008). Muuttujien määrän ollessa vielä pieni laskimme jokaiselle muuttujakombinaa-

Taulukko 1. Aineistona käytetyt aikasarjat ja binäärimuuttujat

Lyhenne	Muuttuja	Tyyppi	Kuvaus
Päämuuttujat (endogeeniset muuttujat)			
M	<i>vol_per_capita</i>	aikasarja	Kasvimaitotuotteiden myyntimäärä grammoissa (g) tai millilitroissa (ml) suhteutettuna Suomen väkilukuun
V^o	<i>veganismi.fo</i>	aikasarja	Veganismiaiheisia sanoja sisältävien foorumiviestien viikoittainen määrä suhteutettuna Futusomen koko foorumiaineistoon
V^{bl}	<i>veganismi.bl</i>	aikasarja	Veganismiaiheisia sanoja sisältävien blogiviestien viikoittainen määrä suhteutettuna Futusomen koko blogiaineistoon
K^o	<i>kasvimaitotuote.fo</i>	aikasarja	Kasvimaitoaiheisia sanoja sisältävien foorumiviestien viikoittainen määrä suhteutettuna Futusomen koko foorumiaineistoon
K^{bl}	<i>kasvimaitotuote.bl</i>	aikasarja	Kasvimaitoaiheisia sanoja sisältävien blogiviestien viikoittainen määrä suhteutettuna Futusomen koko blogiaineistoon
Kontrollimuuttujat (eksogeeniset muuttujat)			
	tuotteet	aikasarja	Myytyjen eri tuotenimikkeiden viikoittainen määrä
	<i>joulu - 1vk</i> <i>joulu</i> <i>joulu + 1vk</i>	binäärinen	Joulupyhien osuminen viikolle. Vuosien 2012 ja 2013 sekä vuosien 2014 ja 2015 joulut mallinnettiin eri tavalla johtuen joulun erilaisesta sijoittumisesta kalenteriviikolle.
	<i>vegaanibaaste</i>	binäärinen	Saa arvon yksi, jos käynnissä on ollut tammikuinen Vegaanihaaste (alkaen vuodesta 2015)

tiolle oman mallin sillä edellytyksellä, että myyntiluvut ja eksogeeniset muuttujat olivat aina osana mallia. Tämä tuotti kaikkiaan 15 VARX-mallia. Kaikille muodostetuille malleille valittiin sopiva viivetermien (p) määrä Akai-ken informaatiokriteerin (AIC) perusteella. Jatkoanalyysiin valittiin harjoitusaineistoon parhaiten sopiva malli, jolla oli korkea korjattu selitysaste (adj. R^2) muuttujan M suhteen lisäk-

si tarkistaen, että mallin virhetermit eivät Portmanteu-testin perusteella osoittaneet merkkejä autokorrelaatiosta.

Paras VARX-malli sisälsi myyntivolyymien suhteutettuna väestöön (M) lisäksi veganismiaiheiset foorumiviestit (V^o) ja kasvimaitoaiheiset blogiviestit (K^{bl}). Tälle muuttujakombinaatiolle optimimäärä viiveitä oli AIC-vertailussa kolme ($p = 3$), eli malli on muotoa VARX(3).

Taulukko 2. Analyysin eteneminen

Vaihe	Kirjallisuus
1. VAR-mallien estimointi	Pecican 2010; Lütkepohl 2009
2. VAR-mallien yhteisintegraatiotesti	Johansen 1988
3. VEC-vektorivirhekorjausmallin muodostaminen	Lütkepohl 2004
4. VEC-mallin analysointi ja diagnostiikka	Lütkepohl 2004
5. VEC-mallin soveltaminen:	Lütkepohl 2005; Wickens ja Motto 2001
a) Ennustaminen	
b) Muuttujien välisten lyhyen aikavälin suhteiden tarkastelu (impulssivasteet)	
6. ARIMA-mallien muodostaminen ja ennustaminen	Box ja Jenkins 1970

Mallin sovittamisen jälkeen testattiin sen diagnostiikat. Valittu VARX-malli ei läpäissyt yhteisintegraatiotestausta ominaisarvoihin perustuvalla Johansenin testillä (Johansen 1988; Juselius 2009), sillä mallin muuttujat paljastuivat yhteisintegroituiksi asteella yksi. Nollahypoteesi $r = 0$ yhteisintegraation puuttumisesta hylättiin arvolla 55,27 ($p < .01$), kun taas nollahypoteesi yhteisintegraation asteesta $r \leq 1$ jäi voimaan arvolla 15,21 (*ns.*). Jos epästationaarisille satunnaismuuttujille löytyy stationaarinen lineaarinen kombinaatio, muuttujat ovat yhteisintegroituja (Murray 1994). Tämän vuoksi siirryimme käyttämään vektorivirhekorjausmallia (*Vector Error Correction Model* VECM, ks. Lütkepohl 2004).

Kaikkien endogeenisten muuttujien malleista tarkastettiin mallien jäännöstermien mahdollinen autokorrelaatio (korrelogrammit ja asymptootinen Portmanteu-testi), niiden normaalijakautuneisuus (kvantiilikuvio ja Shapiro-Wilk-testi) sekä heteroskedastisuus (resi-

duaalien sijoittuminen mallin sovittamien arvojen suhteen). M -muuttujaa selitettäessä jäännöstermit näyttivät korrelogrammin perusteella viitteitä autokorrelaatiosta useammalla viiveellä, mutta läpäisivät kuitenkin muut edellä mainitut diagnostiset tarkastelut. Puolestaan V^{fo} ja K^{bl} -muuttujien jäännöstermit osoittivat visuaalisen tarkastelun perusteella merkkejä heteroskedastisuudesta, eivätkä ne olleet normaalijakautuneita. Kyseisten muuttujien selittäminen ei kuitenkaan ollut tutkimuskysymyksemme kannalta oleellista. Käytimme impulssivasteita (*Impulse Response Function*, ks. Lütkepohl 2005; Wickens ja Motto 2001) mallintamaan standardoidussa impulssimuuttujassa tapahtuvan yhden yksikön suuruisen shokin vaikutusta myyntiin eli vastemuuttujaan. Impulssivasteet keskittyivät muuttujien välisten yhteyksien tarkasteluun.

Vastataksemme ennustamista koskevaan tutkimuskysymykseen testasimme VEC-mallin ennustekykä testiaineiston havainnoilla *vars-*

paketin *predict*-funktiolla, joka pyrkii mallin ja virheenkorjaustermien avulla ennustamaan selitettävän muuttujan arvoja halutulle aika-asteleelle. Koska meitä kiinnosti ennen kaikkea sosiaalisen median viestien käyttö myynnin ennustamisessa, ennusteita tehtiin vain myyntivolyymien kehitykselle. Lisäksi selvitimme (S)ARIMA(X)-mallien avulla sitä, miten kasvimaitotuotteiden myyntiä voidaan ennakoita VEC-malliin valikoituneilla muuttujilla tai toisaalta pelkillä myyntiluvuilla itsellään. ARIMA-mallit tunnistettiin harjoitusaineistolle hyödyntämällä Box-Jenkins-menetelmää, jossa tarkasteltava muuttuja ensin stationarisoitiin. Tämän jälkeen kokeilemalla tunnistettiin mallin viivepolynomien asteluvut, minkä jälkeen mallit estimoitiin ja suoritettiin diagnostiset tarkistukset (Box ja Jenkins 1970). Sosiaalisen median muuttujia sisältävän myynnin SARIMAX-mallin muuttujat valittiin suoraan aiemmin muodostetun VEC-mallin perusteella. Sosiaalisen median muuttujien neljän viikon myynnin rullaaviin ennusteisiin tarvittavat puuttuvat arvot estimoitiin niille muodostetuista ARIMAX-malleista. Viiden ARIMA-mallin ennustekykyä arvioitiin testiaineistolla laskemalla ennusteille keskineliövirheen neliöjuuri (*Root Mean Square Error*, RMSE) toteutuneisiin myyntilukuihin verraten.

Kaikissa rullaavissa ennustemalleissa ennustettiin viimeistä 48 viikon osuutta koko aineistosta ($n = 258$) harjoitusaineiston kokoa kasvattaen ($210 + i$) niin, että ennustettava testiaineiston havainto oli yhden aika-asteleen päässä yhden viikon ennusteille ja neljän päässä neljän viikon ennusteille. Tämä tarkoitti sitä, että yhden viikon ennusteen viimeisen testiaineiston viikon tapauksessa (koko aineiston viikko 258) harjoitusaineistoon kuului 257 viikkoa, kun taas ensimmäistä testiaineiston

viikkoa ennustettaessa harjoitusaineistossa oli 210 viikkoa.

3. Tulokset

3.1 VAR- ja VEC-mallit

VEC(2)-malli estimoitiin VARX(3)-mallin pohjalta. VEC-mallinnuksen tulokset on esitetty taulukoissa 3 ja 4. VEC-mallien tilastollinen kelvollisuus testattiin autokorrelaation, normaaliuden ja heteroskedastisuuden osalta. Raportoimme tässä vain tutkimusongelmamme kannalta oleellisimman mallin eli myyntivolyymia selittävän mallin.

Mallin negatiivinen ECT eli korjaustermi (*error correction term*) kertoo, että pitkän aikavälin riippuvuussuhde mallin muuttujien välillä on olemassa. Taulukossa 3 raportoidut yhteisintegraatiokertoimet kuvaavat mallin muuttujien välisiä suhteita. Taulukossa 4 raportoidut mallin regressiokertoimet puolestaan selittävät lyhyen aikavälin vaikutuksia, eli sitä, millä nopeudella heilahdukset eri muuttujien arvoissa tasoittuvat kohti pitkän aikavälin tasapainoa. Sijoitettuna funktioon pitkän aikavälin malli ilmaistaan seuraavasti:

$$M_t = 1,136V_t^{fo} + 0,273K_t^{bl} + 0,088$$

Veganismi aiheisten foorumiviestien suhteellisten osuuksien muutokset viiveellä $t-2$ siis selittävät positiivisesti muutoksia myyntivolyymissa M ($p = 0,009$). Kasvaimaitotuotteita koskevien blogiviestien K^{bl} vaikutus myyntiin on molemmilla mallin viiveillä ei-merkittävä. Suurimmat kertoimet ja tilastollisesti selkeimmät vaikutukset ovat kuitenkin myyntivolyymilla

Taulukko 3. Yhteisintegraatiovektorien kertoimet (β) ja niiden korjaustermit \hat{a} mallille VEC(2)

	M_t	$V_t^{f_0}$	K_t^{bl}	Vakio
$\hat{\beta}$	1	-1,136 (-9,6)	-0,273 (-2,9)	-0,088 (-1,2)
\hat{a}	-0,085 (-3,1)	0,621 (6,6)	-0,008 (0,2)	

Suluissa t-testisuureet

Taulukko 4. Lyhyen aikavälin korjauskertoimet VEC(2)-mallin endogeeniselle muuttujalle M

Muuttuja	Kerroin	SE	t	p	Tilastollinen merkitsevyys
ect1	-0,085	0,028	-3,089	0,002	**
joulu12_13 ₋₁	1,252	0,152	8,236	<0,001	***
joulu12_13	-2,762	0,177	-15,628	<0,001	***
joulu12_13 ₊₁	0,992	0,240	4,127	<0,001	***
joulu14_15 ₋₁	0,535	0,153	3,491	<0,001	***
joulu14_15	0,191	0,190	1,007	0,315	
joulu14_15 ₊₁	-1,556	0,161	-9,652	<0,001	***
vegaanihaaste	0,705	0,103	6,807	<0,001	***
ΔM_{t-1}	-0,392	0,067	-5,861	<0,001	***
$\Delta V_{t-1}^{f_0}$	0,029	0,021	1,414	0,160	
ΔK_{t-1}^{bl}	-0,008	0,037	-0,211	0,833	
ΔM_{t-2}	-0,304	0,046	-6,651	<0,001	***
$\Delta V_{t-2}^{f_0}$	0,070	0,026	2,642	0,009	**
ΔK_{t-2}^{bl}	0,018	0,037	0,476	0,635	
R^2	0,8318				
R^2 korjattu	0,8196				

Mitsellään sekä eksogeenisellä Vegaanihaaste-muuttujalla (molemmat $p < 0,001$). Koko mallin R^2 selitysaste on 0,8196, mikä voidaan tulkitä varsin kohtuulliseksi.

Impulssivastefunktioita käytettiin mallintamaan yhden muuttujan aiheuttamaa sokkia toiselle muuttujalle. Impulssivasteita varten kaikki tarkasteltavat muuttujat ensin standardisoitiin. Impulssien kohdemuuttujana oli aina myytyjen tuotteiden volyyymi ja impulssimuuttujana myyntivolyyymi itse sekä kaksi VEC-mallissa mukana ollutta sosiaalisen median muuttujaa (kuvio 3). Myyntivolyymin impulssivasteessa (3A) näkyy, miten positiivinen vaikutus myyntimäärään ensin hidastuu ja sitten kiihtyy hiukan ennen vakioitumista tasaiselle vaikutukselle noin viiden viikon jälkeen sokista. Veganismi aiheisten foorumikeskustelujen suhteellinen osuus sen sijaan odotetusti aiheuttaa kasvua myyntivolyyymissä nousevalla vaikutuksella (3B), joka vaikuttaisi terävän alkuefektin jälkeen olevan myös pitkäaikainen. Kasvimaitotuotteita koskevien blogikirjoitusten impulssivasteen keskiarvo (3C) on myös positiivinen, mutta marginaalinen, vaikutuksen luottamusvälin alarajan ollessa negatiivinen. Nämä havainnot ovat linjassa VEC(2)-mallin kertoimien kanssa.

3.2 Myynnin ennustaminen VEC- ja ARIMA-malleilla

ARIMA-mallit, joilla selvitetiin kasvimaitotuotteiden myyntilukujen aikasarjan omaa ennustekykyä, on esitelty taulukossa 5. Malleihin päädyttiin sekä itse kokeilemalla että hyödyntämällä koneistetusti *forecast* R-kirjaston *auto.arima*-funktioita (Hyndman ym. 2018). Parhaat ARIMA-mallit on listattu tyypeittäin. Kaikissa malleissa sarja on differentioitu stationaarisuu-

den saavuttamiseksi. I(1)-myyntisarjan stationaarisuus testattiin ADF- ja KPSS-testeillä (ks. liite 1). Taulukon viimeisissä SARIMAX-malleissa on huomioitu kausivaihtelu. Kauden pituus viikkoaineistolla on 52. Mallit, jotka ovat muotoa (S)ARIMAX sisältävät eksogeeniset binäärimuuttujat (ks. taulukko 1). Yksikään taulukon 5 malli ei kuitenkaan selitä väestöön suhteutettua myyntiä tyydyttävästi, sillä mallien virhetermit ovat Ljung-Box-testin perusteella autokorreloituneita. Mallien ARIMA-termien kertoimia ei raportoida, koska niihin ei voida luottaa jäännöstermien autokorrelaation vuoksi. Malleja voidaan kuitenkin hyödyntää ennustamiseen. Taulukon malleista ennustekyvyltään parhaaksi RMSE:n perusteella osoittautui vain myyntiaineiston huomioiva SARIMAX(2,1,0) (1,0,0)₅₂-malli sekä yhden että neljän viikon testiaineistolle tehdyissä rullaavissa ennusteissa. VEC-mallinnukseen pohjaavassa ennusteessa RMSE yhden viikon rullaavalle ennusteelle oli 1,323 ja neljän viikon ennusteelle 1,699. Näin ollen molemmat ennustemallit häviävät (S)ARIMAX-ennusteille.

4. Johtopäätökset

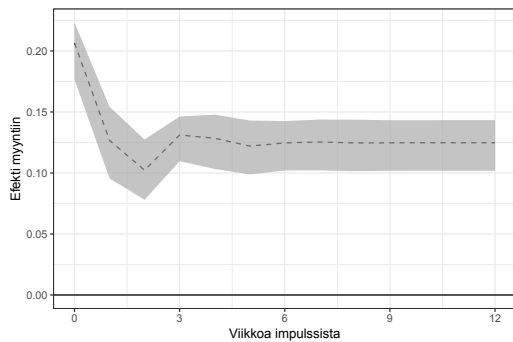
4.1 Tutkimuksen johtopäätökset

Aineistomme aikasarjat osoittavat, että kiinnostus veganismia ja vegaanisia kasvimaitotuotteita kohtaan on kasvanut sekä tuotteiden myyntivolyyymillä että sosiaalisen median normalisoidulla viestivolyyymillä mitattuna. Ilmiön nousu näkyy erityisesti aineiston viimeisenä vuonna 2016. Ekonometrinen mallinnuksemme osoitti, että aikasarjojen välillä on merkitseviä yhteyksiä.

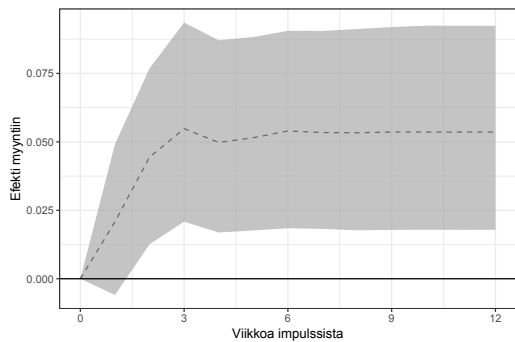
ARIMA- ja VEC-mallinnukset osoittivat, että selitettäessä kasvimaitotuotteiden myynti-

Kuvio 3. Myyntivolyymin impulssivasteet muuttujille M , V^{fo} ja K^{bl}

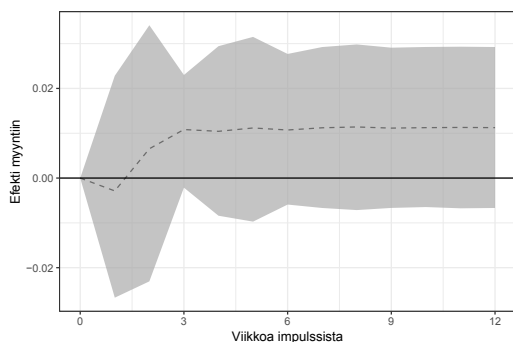
3A. Myyntivolyymin (M_t) impulssivaste



3B. Veganismiaiheisten foorumiviestien määrän V^{fo} impulssivaste



3C. Kasvimaitoaiheisten blogiviestien määrän K^{bl} impulssivaste



95 % luottamusvälit estimoitu bootstrap-menetelmällä ($n = 400$)

määrää sen historiallisilla arvoilla, olennaista on tarkastella muutamaa edeltävää viikkoa. VEC(2)-mallin myynnin selitysmallissa kahden edellisen viikon myynnin muutokset ovat lyhyellä aikavälillä negatiivisessa yhteydessä annetun viikon t myynnin muutokseen. Tuloksemme täten heijastanee ruokaostoksiin liittyvää luonnollista, heilurimaista sykliä. Kirjallisuuden mukaan kuluttajien ruokakaupassa käymisen tiheys riippuu ostoksilla käymiseen

liittyvistä matkakustannuksista ja toisaalta ruokien säilyttämiseen liittyvistä varastonpito-kustannuksista (Bawa ja Ghosh 1999). Kotitaloudet siis pyrkivät löytämään tasapainon sen suhteen, kuinka usein ruokakaupassa käydään, eivätkä sen takia välittömästi reagoi yksittäisiin tuotteisiin liittyviin kulutusärsykkeisiin kaupan ulkopuolella. Kotitaloudet tyypillisesti suosivatkin harvempia mutta suuria ostosker-toja (esim. kerran viikossa tai kahdessa viikos-

Taulukko 5. ARIMA-mallien sovitus ja ennustekyky

Malli	RMSE harjoitus- aineistolle (n=210)	ARIMA- jäännöstermit autokorreloituneet (Ljung-Box-testi)	Testiaineistolle tehtyjen yhden viikon rullaavien ennusteiden RMSE (48 erillistä ennustetta 48 havainnolle)	Testiaineistolle tehtyjen neljän viikon rullaavien ennusteiden RMSE (12 erillistä ennustetta 48 havainnolle)
ARIMA(0,1,1)	1,796	Kyllä	1,720	2,335
ARIMAX(0,1,1)	0,880	Kyllä	1,275	1,722
ARIMAX(0,1,1) kahdella fourier-termillä	0,816	Kyllä	1,261	1,604
SARIMAX(2,1,0)(1,0,0) ₅₂	0,807	Kyllä	1,183	1,459
SARIMAX(2,1,0)(1,0,0) ₅₂ sosiaalisen median viive- muuttujilla (V^{jo} , K^{bl})	0,802	Kyllä	1,193	1,508

Sosiaalisen median muuttujat V^{jo} ja K^{bl} ovat viiveillä 2 ja 3.

sa), joita täydennetään pienillä ostoksilla lähikaupoista (Dibsdall ym. 2003; Yoo ym. 2006). Impulssivastineita tarkastelemalla voidaan kuitenkin todeta, että yksittäisellä myyntisokilla on pitkän aikavälin myyntiä nostava vaikutus. Lisäksi SARIMAX(2,1,0)(1,0,0)₅₂-mallin hyvästä ennustekyvystä voidaan todeta, että myynnin vuosittaisilla kausiefekteillä on selvästi merkitystä kasvimaitotuotteiden myyntiä ennakoivissa.

VEC-mallien negatiivisen korjaustermin (ECT) perusteella pitkän aikavälin yhteys myyntivolyymiin ja sosiaalisen median viestien määrän välillä on olemassa. Tämän lisäksi mal-

li antaa tietoa lyhyen aikavälin dynamiikasta. VAR-mallinnuksen perusteella selittäviksi muuttujiksi valikoituivat veganismia koskevat foorumiviestit sekä kasvimaitoja koskevat blogiviestit. Kun selitetään kasvimaitotuotteiden myyntiä sosiaalisen median muuttujilla, veganismia koskevien foorumiviestien määrällä havaittiin olevan positiivinen yhteys myyntiin. Kasvimaitotuotteita koskevien blogiviestien määrällä ei sen sijaan ollut merkitsevää yhteyttä myyntiin. Analyysimme siis osoittaa, että yleistä ilmiötä koskevat keskustelut ovat tuotekategorian myyntiä paremmin selittävä muuttuja kuin tuotteita koskevat keskustelut. Tulos-

ta voidaan tulkita esimerkiksi siten, että blogiviestit, joissa mainitaan yksittäisiä kasvimaitotuotteita (esim. kauramaito, soijamaito), ovat tyypillisimmin ruokaohjeita ja reseptejä. Foorumikeskustelut sen sijaan oletettavasti heijastelevat suoraviivaisemmin kuluttajien ruokavalintoja ja niistä käytäviä ajankohtaisia keskusteluja (ks. Pantzar ja Ruckenstein 2018).

Veganismiaiheiset foorumiviestit selittävät kasvimaitotuotteiden myynnin muutoksia kahden viikon viiveellä. Foorumien suhteellisten viestimäärien muutokset ennakoivat siis muutoksia myyntivolyymissa. Emme kuitenkaan väitä, että foorumiviestit yksinään vaikuttaisivat kuluttajien ostopäätöksiin, vaan ne ennemmin kuvaavat kuluttajien kiinnostusta veganismia kohtaan. Onkin oletettavaa, että sosiaalisen median vaikutus on kuluttajien kohdalla kumulatiivista ja osa laajempaa kertyneiden tietojen ja tunteiden kokonaisuutta; sosiaalisen median viestit indikoivat olemassa olevaa kiinnostusta ja mahdollisia ostoaikeita. Lisäksi sosiaalisen median viestit vaikuttavat niitä lukeviin kuluttajiin; niiden on aiemmin todettu olevan vaikuttava tietolähde ja vaikutin kuluttajien ostopäätöksille (Bickart ja Schindler 2001; Kim ja Hanssens 2017; Stephen ja Galak 2012). Tämä on kuitenkin tietääksemme ensimmäinen tutkimus, jossa analysoidaan yhteyttä ilmiötason viestien ja kategoriataason myynnin välillä. Tulostemme mukaan myynnin näkökulmasta on tärkeää, että uudet tuotteet kytkeytyvät johonkin laajempaan kuluttajia kiinnostavaan ilmiöön.

Analyysin mukaan myös Vegaanihaasteella on merkitsevä suhde kasvimaitotuotteiden myyntiin. Näyttäisi siis siltä, että tämä kymmeniä tuhansia ihmisiä kiinnostava haaste (Helsingin Uutiset 2018) vaikuttaa kuluttajien ostopäätöksiin siinä määrin, että se selittää kasvi-

maitotuotteiden myyntimäärää tammikuussa. Toki on muistettava, että tammikuussa monilla on tapana muutenkin pyrkiä terveellisiin tai hyveellisiin elämäntapoihin (Pölkki 2019), mikä saattaa näkyä myös kasvimaitotuotteiden lisääntyneenä kuluttamisena.

Ennustemalleista parhaiten toimii SARIMAX(2,1,0)(1,0,0)₅₂-malli, joka huomioi ainoastaan myyntivolyymin muutokset sekä kontrollimuuttujat. Vaikka sosiaalisen median keskusteluaineistoja sisältävä VEC(2)-malli osoitti, että tutkituilla aikasarjoilla on pitkän aikavälin riippuvuusuhde, sekä VEC(2)-malliin pohjaava ennuste että sosiaalisen median muuttujia sisältävä SARIMAX-malli toimivat ennustemalleina heikommin. Tähän voi olla lukuisia eri syitä. Yksi selitys voi olla, että sosiaalisen median viestien vaikutukset ovat niin pieniä, että ennustemallina pelkällä myyntidatalla ennustaminen toimii paremmin ja varmemmin. Eräs merkittävä ennustetta hankaloittava tekijä voi olla myös veganismin tunnettuuden selkeä kasvu aineiston loppua kohden: ilmiön kasvu näkyy silmämääräisesti sekä myyntiluvuissa että sosiaalisen median viestien suhteellisissa osuuksissa (kuviot 1 ja 2). Siten sarjan testiaineistossa tapahtuu mahdollisesti rakenteellinen murros, joka hankaloittaa ennustamista, ennustemallien rakenteiden määräytyessä murrosta edeltävillä harjoitusaineiston havainnoilla.

4.2 Rajoitukset ja tulevaisuuden tutkimusaiheet

Tutkimuksesta on syytä huomioida muutamia rajoituksia ja tunnistaa jatkotutkimusmahdollisuuksia. Aineistomme rajautuu viiteen vuoteen (2012–2016), minkä johdosta kasvimaitotuotteiden myynti näyttäytyy kasvavana ilmiö-

nä. Jos olisimme valinneet tarkasteluun aikaisemman ajankohdan, jolloin tutkitut tuotteet olivat vähemmän trendikkäitä, sosiaalisen median viestien määrä ja vaihtelu olisivat todennäköisesti olleet hyvin erilaisia. Esimerkiksi ei-trendikkään punaisen lihan myynnin mallintaminen sosiaalisen median aineiston pohjalta olisi saattanut tuottaa hyvin erilaisia tuloksia. Aiemmin onkin huomattu, että kuluttajien tuottama ruokaan liittyvä sosiaalisen median sisältö painottuu vain tietynlaisiin ruokiin (Abbar ym. 2015; De Choudhury ym. 2016). On myös syytä muistaa, ettei VECM ole teorialähtöinen analyysimenetelmä (Pecican 2010), joten tulokset eivät paljasta, mistä havaitut tulokset johtuvat. Jatkotutkimusta tarvitaankin selvittämään sosiaalisen median keskustelujen ja kuluttajatuotteiden kysynnän välisiä vaikutusmekanismeja.

Isoihin data-aineistoihin ja niiden ennakoitavoimaan liittyvää hybristä on myös aiheellisesti kritisoitu erityisesti sen jälkeen, kun Google Flu Trends -ennustemalli osoittautui virheelliseksi (Lazer ym. 2014) ja sosiaalisen median ennustevoima vaalituloksiin liioitelluksi (Gayo-Avello 2013). Lazer ym. (2014) muistuttavatkin, että myös suuria massadatoja käytettäessä on pohdittava aineiston ja mitaustapojen validiteettia ja toisaalta otettava huomioon tutkittujen ilmiöiden sosioteknistä dynamiikkaa. Sosiaalisen median datan luonteeseen vaikuttavat esimerkiksi palvelujen algoritmit ja perinteisen uutismedian tuottamat

sisällöt. Lukijan onkin hyvä muistaa, että sosiaalinen media on hyvin moninainen joukko erilaisia, eri tavoin käytettyjä ja ajan yli muuttuvia alustoja. Esimerkiksi aineistomme perusteella blogikirjoitusten absoluuttinen määrä on kääntynyt tutkittuna aikajaksona selvästi laskuun. Täten muutokset alustojen keskinäisessä suosiossa ja toiminnallisuuksissa rajoittavat tutkimustuloksiamme. Lisäksi hakusanapohjaisessa aineiston rajauksessa tulee huomioida muutokset ihmisten käyttämässä arkikielessä ja sen mahdollisessa vaikutuksessa löydettyjen viestien määrään. Nykyinen operationalisointimme vastaa kuitenkin käytännössä yleisesti käytössä olevia mediaseurannan avainsanakuihin pohjautuvia menetelmiä, jotka voisivat olla markkina-analyttikkojen käytössä.

Sosiaalinen media mahdollistaa kuluttamiseen liittyvien teemojen kiinnostuksen mittaamisen ajassa ja paikassa sekä suurimittaisen, ei-tunkeilevan aineiston keräämisen. Tuloksemme antavat viitteitä siitä, että tällaisilla aineistoilla on potentiaalisesti selitysvoimaa kuluttajatrendien ennakoinnissa, mutta parhaimman mallinnustavan kartoittaminen vaatii jatkotutkimuksia. Ehdotamme, että jatkotutkimuksessa otettaisiin tarkemmin huomioon esimerkiksi sosiaalisen median viestien julkaisukonteksti ja niiden saavuttamien yleisöjen laajuus. Tällöin voidaan entistä tarkemmin tunnistaa, minkälaiset sosiaalisen median sisällöt vaikuttavat kuluttajien ostopäätöksiin. □

Kirjallisuus

- Abbar, S., Mejova, Y. ja Weber, I. (2015), "You tweet what you eat: Studying food consumption through Twitter", *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, The Association for Computing Machinery (ACM).
- Aramaki, E., Maskawa, S. ja Morita, M. (2011), "Twitter catches the flu: detecting influenza epidemics using Twitter", *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics: 1568–1576.
- Asur, S. ja Huberman, B. A. (2010), "Predicting the future with social media", *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, IEEE Computer Society: 492–499.
- Bawa, K. ja Ghosh, A. (1999), "A model of household grocery shopping behavior", *Marketing Letters* 10: 149–160.
- Bickart, B. ja Schindler, R. M. (2001), "Internet forums as influential sources of consumer information", *Journal of Interactive Marketing* 15: 31–40.
- Bittman, M. (2013), VB6: *Eat Vegan before 6:00 to Lose Weight and Restore your Health... for Good*, Clarkson Potter/Publishers.
- Bollen, J., Mao, H. ja Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science* 2: 1–8.
- Box, G. ja Jenkins, G. (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day.
- Brandt, P. ja Williams, J. (2007), *Multiple Time Series Models*, SAGE Publications.
- Chintagunta, P. K., Gopinath, S. ja Venkataraman, S. (2010), "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets", *Marketing Science* 29: 944–957.
- Culotta, A. (2010), "Towards detecting influenza epidemics by analyzing Twitter messages", *Proceedings of the First Workshop on Social Media Analytics*, The Association for Computing Machinery (ACM): 115–122.
- De Choudhury, M., Sharma, S. ja Kiciman, E. (2016), "Characterizing dietary choices, nutrition, and language in food deserts via social media", *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, The Association for Computing Machinery (ACM): 1157–1170.
- Dellarocas, C., Zhang, X. M. ja Awad, N. F. (2007), "Exploring the value of online product reviews in forecasting sales: The case of motion pictures", *Journal of Interactive Marketing* 21: 23–45.
- Dewan, S. ja Ramaprasad, J. (2014), "Social media, traditional media, and music sales", *MIS Quarterly* 38: 101–121.
- Dhar, V., ja Chang, E. A. (2009), "Does chatter matter? The impact of user-generated content on music sales", *Journal of Interactive Marketing* 23: 300–307.
- Dibsdall, L. A., Lambert, Nn, Bobbin, R. F. ja Frewer, L. J. (2003), "Low-income consumers' attitudes and behaviour towards access, availability and motivation to eat fruit and vegetables", *Public Health Nutrition* 6: 159–68.
- DiGrazia, J., McKelvey, K., Bollen, J. ja Rojas, F. (2013), "More tweets, more votes: Social media as a quantitative indicator of political behavior", *PloS ONE* 8.
- Djurens Rätt (2017), "Opinionsundersökning, Våren 2017", Novus 5.5.2017, <https://www.djurensratt.se/sites/default/files/2017-06/vegoopinion-novus-2017.pdf> (haettu 11.1.2019).
- Djurens Rätt. (2015), *Var tionde Svensk väljer fortfarande vego*, blogikirjoitus, <https://www.djurensratt.se/blogg/var-tionde-svensk-valjer-fortfarande-vego> (haettu 23.1.2019).

- Francione, G. (2010), *The Animal Rights Debate: Abolition or Regulation*, Columbia University Press.
- Fried, D., Surdeanu, M., Kobourov, S., Hingle, M. ja Bell, D. (2014), “Analyzing the language of food on social media”, *2014 IEEE International Conference on Big Data*, Institute of Electrical and Electronics Engineers (IEEE): 778–783.
- Gayo-Avello, D. (2013), “A meta-analysis of state-of-the-art electoral prediction from Twitter data”, *Social Science Computer Review* 31: 649–679.
- Gencoglu, O. ja Ermes, M. (2018), “Predicting the flu from Instagram”, arXiv preprint, <https://arxiv.org/abs/1811.10949> (haettu 11.1.2019).
- Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolin-ski, M. ja Brilliant, L. (2009), “Detecting Influenza Epidemics Using Search Engine Query Data”, *Nature* 457: 1012–1014.
- Hancox, D. (2018), “The unstoppable rise of veganism: how a fringe movement went mainstream”, *The Guardian* 1.4.2018.
- Helsingin Uutiset (2018), ”Vegaanihaasteella hurja suosio – vanha ennätys ylitettiin jo kaksinkertaisesti”, *Helsingin Uutiset* 10.1.2018.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O’Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. ja Yasmeeen, F. (2018), forecast: *Forecasting functions for time series and linear models*, R package version 8.4, <http://pkg.robjhyndman.com/forecast>.
- Jallinoja P., Vinnari, M. ja Niva, M. (2019a), “Ve-ganism and plant-based eating: analysis of interplay between discursive strategies and lifestyle political consumerism”, teoksessa Boström, M., Micheletti, M., Oosterveer, P. (toim.), *The Oxford Handbook of Political Consumerism*, Oxford University Press, Oxford.
- Jallinoja, P., Jauho, M. ja Pöyry, E. (2019b), “Miten Suomi söi 2008–2016? Erityisruokavaliot ja niiden taustatekijät”, *Yhteiskuntapolitiikka*.
- Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12: 231–254.
- Jones, A., Omer, S., Bednarczyk, R., Halsey, N., Moulton, L. ja Salmon, D. (2012), “Parents’ source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions”, *Advances in Preventive Medicine*: 1–8.
- Jungherr, A., Jürgens, P. ja Schoen, H. (2012), “Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. ‘Predicting elections with Twitter: What 140 characters reveal about political sentiment’”, *Social Science Computer Review* 30: 229–234.
- Juselius, K. (2009). “The long swings puzzle: what the data tell when allowed to speak freely”, teoksessa Mills, T. ja Patterson, K. (toim.), *Palgrave Handbook of Econometrics*, Palgrave Macmillan: 349–384.
- Kallionpää, K. (2016), ”Vegaanihaaste villitsee, ja ravitsemussuosituksia päivitetään – asiantuntijat vastaavat, miten kasvisruokavalio vaikuttaa terveyteen”, *Helsingin Sanomat* 8.1.2016.
- Kerschke-Risch, P. (2015), “Vegan diet: motives, approach and duration. Initial results of a quantitative sociological study”, *Ernährungs Umschau* 62: 98–103.
- Kesko (2017), ”Kasvipohjaisten tuotteiden myynti kasvaa edelleen voimakkaasti – vegehilly löytyy jo 200 K-ruokakaupasta”, Kesko Oy, Lehdistötiedote 9.6.2017.
- Kim, H. ja Hanssens, D. (2017), “Advertising and Word-of-Mouth Effects on Pre-launch Consumer Interest and Initial Sales of Experience Products”, *Journal of Interactive Marketing* 37: 57–74.

- Laaksonen, S.-M., Matikainen, J. & Tikka, M. (2013), Tutkimusotteita verkosta, Julkaisussa: Laaksonen, S.-M., Matikainen, J. & Tikka, M. (toim.), *Otteita verkosta, Verkon ja sosiaalisen median tutkimusmenetelmät*, Tampere: Vastapaino.
- Laaksonen, S.-M. ja Pöyry, E. (2018), “Pahastumista vai politiikkaa: Sosiaalisen median kohut affektiivisessä keskustelukulutturissa”, *WiderScreen* 11: 1–37.
- Lazer, D., Kennedy, R., King, G., ja Vespignani, A. (2014), “The parable of google flu: Traps in big data analysis”, *Science* 343(6176): 1203–1205. <https://doi.org/10.1126/science.1248506>
- Luo, X., Zhang, J. ja Duan, W. (2013), “Social media and firm equity value”, *Information Systems Research* 24: 146–163.
- Lütkepohl, H. (2009), “Econometric analysis with vector autoregressive models”, *Handbook of Computational Econometrics*, John Wiley & Sons: 281–319.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.
- Lütkepohl, H. (2004), “Vector autoregressive and vector error correction models”, teoksessa Lütkepohl H. ja Kräzig M. (toim.), *Applied Time Series Econometrics*, Cambridge University Press: 86–158.
- Mayr, P. ja Weller, K. (2017), “Think before you collect: setting up a data collection approach for social media studies”, teoksessa Sloan, L. ja Quan-Haase, A. (toim.), *The SAGE Handbook of Social Media Research Methods*, SAGE Publications: 107–124.
- Meindertsma, C. (2008), *PIG 05049*, Flocks.
- Moon, S., Bergey, P. ja Iacobucci, D. (2010), “Dynamic effects among movie ratings, movie revenues, and viewer satisfaction”, *Journal of Marketing* 74: 108–121.
- Morris, C. (2018), “‘Taking the politics out of broccoli’: Debating (de)meatification in UK national and regional newspaper coverage of the Meat Free Mondays campaign”, *Sociologia Ruralis* 58: 433–452.
- Murray, M. (1994), “A Drunk and Her Dog: An Illustration of Cointegration and Error Correction”, *The American Statistician* 48: 37–39
- Murthy, D. (2015), “Twitter and elections: are tweets, predictive, reactive, or a form of buzz?”, *Information Communication & Society* 18: 816–831.
- Nelimarkka, M. ja Laaksonen, S.-M. (2018), “Bitit ja politiikka: tervetuloa laskennallinen yhteiskuntatieteen tutkimus”, *Politiikka* 60: 130–131.
- O’Connor, A. J. (2013), “The power of popularity: An empirical study of the relationship between social media fan counts and brand company stock prices”, *Social Science Computer Review* 31: 229–35.
- Onishi, H., ja Manchanda, P. (2012), “Marketing activity, blogging and sales”, *International Journal of Research in Marketing* 29: 221–234.
- Pantzar, M. ja Ruckenstein, M (2018), “Verkkokeskustelut: Riidankylvämistä ja rauhanrakentamista”, teoksessa Autio, A., Autio, M., Kylkilähti, E. ja Pantzar, M. (toim.), *Kulutus ja talous: Näkökulmia yhteiskunnan muutokseen*, Helsingin yliopisto, Department of Economics and Management Publications No. 70.
- Pecican, E. (2010), “Forecasting based on open VAR model”, *Journal for Economic Forecasting, Institute for Economic Forecasting* 13: 59–69.
- Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software* 27: 1–32.
- Päivittäistavarakauppa ry (2013), “Päivittäistavarakaupan myyntitiedot ja markkinaosuudet 2012”, <https://www.pty.fi/ajankohtaista/tiedotteet/uutinen/article/paeivittaeistavarakaupan-myyntitiedot-ja-markkinaosuudet-2012/> (haettu 29.1.2019).

- Päivittäistavarakauppa ry (2016), "Päivittäistavarakaupan tilastot", <https://www.ptv.fi/julkaisut/tilastot/> (haettu 29.1.2019).
- Pölkki, M. (2019), "Mitä jos Suomi söisi koko vuoden kuten tammikuussa? Muutos kansanterveyteen olisi mullistava, ellei ryhtiliike lopah-taisi jo helmikuussa", *Helsingin Sanomat* 19.1.2019.
- Pöyry, E., Laaksonen, S.-M., Kekkonen, A., ja Pääk-
könen, J. (2018), "Anatomy of Viral Social Me-
dia Events", Proceedings of the 51st Hawaii In-
ternational Conference on System Sciences:
2173–2182. <https://scholarspace.manoa.hawaii.edu/bitstream/10125/50160/1/paper0273.pdf>
- Sharma, S. ja De Choudhury, M. (2015), "Measuring and characterizing nutritional information of food and ingestion content in instagram", *Proceedings of the 24th International Conference on World Wide Web*, The Association for Computing Machinery (ACM): 115–116.
- Stephen, A. T. ja Galak, J. (2012), "The effects of traditional and social earned media on sales: A study of a microlending marketplace", *Journal of Marketing Research* 49: 624–639.
- SVT Suomen virallinen tilasto (2018), "Väestön tieto- ja viestintätekniikan käyttö [verkkojulkaisu]", ISSN=2341–8699, Helsinki: Tilastokeskus, <http://www.stat.fi/til/sutivi/index.html> (haettu 23.1.2019).
- The Vegan Society (2018), "Veganism in the UK", <https://www.vegansociety.com/news/media/statistics> (haettu 23.1.2019).
- Tuhkuri, J. (2016), "ETLANow: A Model for Forecasting with Big Data – Forecasting Unemployment with Google Searches in Europe", *ETLA Reports* 54.
- Tuhkuri, J. (2014), "Big Data: Google Searches Predict Unemployment in Finland", *ETLA Reports* 31.
- Tumasjan, A., Sprenger, T., Sandner, P., & Welp, I. (2010), "Predicting elections with Twitter: What 140 characters reveal about political sentiment", *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence (AAAI): 178–185.
- Véron, O. (2016), "From Seitan Bourguignon to Tofu Blanquette: Popularizing Veganism in France with Food Blogs", teoksessa Castricano, J., Simonsen, R. (toim.), *Critical Perspectives on Veganism*, Palgrave Macmillan, London.
- Whipp, L. (2016), "Big business identifies appetite for plant-based milk". *Financial Times* 15.7.2016. <https://www.ft.com/content/7df72c04-491a-11e6-8d68-72e9211e86ab>
- Wickens, M., & Motto, R. (2001), "Estimating Shocks and Impulse Response Functions", *Journal of Applied Econometrics* 16: 371–387.
- Wilson, K. ja Keelan, J. (2013), "Social media and the empowering of opponents of medical technologies: the case of anti-vaccinationism", *Journal of Medical Internet Research* 15 (e103): 1–4.
- Yang, S. Y., Mo, S. Y. K., & Liu, A. (2015), "Twitter financial community sentiment and its predictive relationship to stock market movement", *Quantitative Finance*, 15(10): 1637–1656.
- Yasseri, T., ja Bright, J. (2014), "Can electoral popularity be predicted using socially generated big data?", *Information Technology*, 56: 246–253.
- Yoo, S., Baranowski, T., Missaghian, M., Baranowski, J., Cullen, K., Fisher, J. O., Watson, K., Zak-
eri, I. F. ja Nicklas, T. (2006), "Food-purchasing patterns for home: a grocery store-intercept survey", *Public Health Nutrition* 9: 384–393.
- Yu, Y., Duan, W. ja Cao, Q. (2013), "The impact of social and conventional media on firm equity value: A sentiment analysis approach", *Decision Support Systems* 55: 919–926.

Liitetaulukot

Liitetaulukko 1. Pyöristetyt yksikköjuuri- ja stationaarisuustestien p-arvot kerran diffensoiduille aikasarjoille

	Augmented Dickey–Fuller (ADF)	Kwiatkowski–Phillips–Schmidt–Shin (KPSS)
vol_per_capita	0,01	0,10
veganismi.fo	0,01	0,10
veganismi.bl	0,01	0,10
kasvimaidot.fo	0,01	0,10
kasvimaidot.bl	0,01	0,10

Liitetaulukko 2. Johansenin testit ominaisvektoreilla

	testisuure	10 pt	5 pt	1 pt
$r = 2$	1,39	7,52	9,24	12,97
$r = 1$	15,21	13,75	15,67	20,20
$r = 0$	55,27	19,77	22,00	26,81