

OSAVUOSIKATSAUKSEN SÄVY JA OSAKKEEN ARVO

Jyväskylän yliopisto
Kauppakorkeakoulu

Pro gradu -tutkielma

2022

Tekijä: Juuso Salminen
Oppiaine: Laskentatoimi
Ohjaaja: Antti Rautiainen



JYVÄSKYLÄN YLIOPISTO

TIIVISTELMÄ

Tekijä Juuso Salminen	
Työn nimi Osavuositarkastuksen sävy ja osakkeen arvo	
Oppiaine Laskentatoimi	Työn laji Pro gradu -tutkielma
Aika (pvm.) 16.3.2022	Sivumäärä 60
<p>Tiivistelmä</p> <p>Talousraportoinnissa esiintyvien laskentatoimen tunnuslukujen merkitystä sijoittajan päätöksenteon kannalta on jo pitkään tutkittu arvorelevanssitutkimuksen muodossa, mutta raporttien tekstimuotoisen sisällön arvorelevanssi on vähäisemmälle huomiolle jäänyt tutkimuskohde. Raporttien teksti saattaa sisältää sijoittajan kannalta kiinnostavia mainintoja esimerkiksi kasvusta, näkymistä tai muusta vaikeasti kvantifioitavasta aiheesta.</p> <p>Tässä tutkimuksessa tarkastellaan talousraporttien tekstisisältöä niiden sävyä mittaamalla. Raporteista määritetään positiivisuus, negatiivisuus ja epävarmuus laskemalla näihin kategorioihin luokiteltujen sanojen prosenttiosuuksia koko raportin sanamäärästä. Sanojen luokitteluun käytetään Loughranin ja McDonaldin (2011) koostamaa sanalista, joka sopii taloudellisten sanojen sävyn tunnistamiseen. Sävyanalyysi suoritetaan suomalaisten pörssiyritysten vuosina 2017–2021 julkaisemille osavuositarkastuksille, joiden julkistuksen yhteydessä syntyneitä epänormaaleja tuottoja selitetään sävymuuttujilla. Toisena tarkastelun kohteena tässä tutkimuksessa ovat positiivisten sanojen lähellä esiintyvät sanat, jotka ovat yhteydessä eroavaan epänormaaliin tuottoon, kun verrataan esiintymistä tilanteeseen, jossa positiivisen sanan ja toisen sanan yhdistelmää ei esiinny raportissa.</p> <p>Tuloksista on huomattavissa, että raportin negatiivisuus pystyy selittämään epänormaaleja tuottoja tilastollisesti merkitsevällä tasolla, kun muutamat yrityksen taloudellista tilaa kuvaavat tunnusluvut ovat vakioituja. Negatiivisuuden kasvaessa epänormaalit tuotot näyttäisivät laskevan, mikä on linjassa aikaisemman tutkimuksen kanssa. Positiivisuus ja epävarmuus eivät näyttäneet yhtä varmoina osakekurssin vaihteluiden selittäjinä. Lähisanatarkastelussa nousi esiin muutama selkeä ryhmä, kuten kasvuun, toimintaan tai aineetomaan pääomaan liittyvät sanat, jotka positiivisen sanan lähistöllä esiintyessään olivat yhteydessä keskimääräisten epänormaalien tuottojen tilastollisesti merkitsevään nousuun tai laskuun.</p>	
Asiasanat talousraportointi, raportoinnin sävy, tekstianalyysi, epänormaalit tuotot	
Säilytyspaikka Jyväskylän yliopiston kirjasto	

SISÄLLYS

TIIVISTELMÄ	2
1 JOHDANTO.....	5
2 TALOUSRAPORTOINTI	9
2.1 Yleistä	9
2.1.1 Talousraportoinnin merkitys sijoittajan kannalta	10
2.1.2 Osavuositarkastus.....	11
2.2 Sanallinen informaatio raportoinnissa	14
3 TEKSTIANALYYSI LASKENTATOIMESSA.....	19
3.1 Tekstianalyysi menetelmänä.....	19
3.2 Raportin sisältö ja osakkeen arvo	21
3.3 Muut kontekstit.....	23
4 AINEISTO JA MENETELMÄT	26
4.1 Aineisto	26
4.2 Menetelmät	28
4.2.1 Tapahtumatutkimus	29
4.2.2 Tekstianalyysi	31
4.2.3 Muuttujat ja regressiomallit.....	32
4.2.4 Positiivisten sanojen lähisanat	36
5 TULOKSET.....	38
5.1 Regressioanalyysi	38
5.2 Lähisanat	42
6 JOHTOPÄÄTÖKSET	46
LÄHTEET	50
LIITE 1 Osavuositarkastukset yhtiöittäin (1 / 2)	57
LIITE 2 Osavuositarkastukset yhtiöittäin (2 / 2)	58
LIITE 3 Esimerkkejä sävysanoista	59
LIITE 4 Koodiesimerkkejä	59

KUVIOT JA TAULUKOT

Kuvio 1 Sävy­muuttujien keskiarvokäyrät aineistossa	27
Taulukko 1 Aineiston koko eri vaiheissa	27
Taulukko 2 Regressiomalleissa käytetyt muuttujat.....	33
Taulukko 3 Aineiston ominaisuuksia.	34
Taulukko 4 Muuttujien välisiä korrelaatioita.....	34
Taulukko 5 Regressiomallin 1 tulokset.	39
Taulukko 6 Regressiomallin 2 tulokset	40
Taulukko 7 Regressiomallin 3 tulokset	41
Taulukko 8 Regressiomalli 1 eri ajanjaksoilla.....	42
Taulukko 9 Positiivisten sanojen lähisanat.....	43

1 JOHDANTO

Nykyisin sijoittajalla on saatavillaan paljon eri informaatiolähteitä sijoituspäätöksiensä tueksi. Eräs niistä on perinteinen tilinpäätös- tai osavuosisiraportointi, jossa tieto perustuu yrityksen omaan raportointijärjestelmään. Vaikka huomio raporteissa yleensä keskittyy niiden tarjoamaan numeraaliseen tietoon yrityksen tilasta, ne sisältävät myös paljon sanallista informaatiota. Raporttien taulukko-osuuksien numerot ovat pohjimmiltaan menneisyyteen pohjautuvia, mutta tekstiosuuksissa saattaa esiintyä myös toimintaa selittävää tai ennakkoivaa tietoa, mikä voi laajentaa raportin tarjoamaa informaationsisältöä (ks. esim. Li, 2006). Tämä tutkielman tarkoituksena on tutkia, voiko raporttien sanallisen osuuden sävyn avulla selittää lyhyen aikavälin osaketuottoja. Tutkimuksen kohteena ovat suomalaisten pörssiyritysten julkaisemat osavuosisiraportit vuosina 2017–2021. Raporttien sävy mitataan jakamalla niissä esiintyvät sanat sävykategorioihin Loughranin ja McDonaldin (2020) sanalistan perusteella. Sävykategorioiden suhteellisista suuruuksista raportin sanamäärään nähden voidaan päätellä raportin sävyn laatu, esimerkiksi suuri prosenttiosuus positiiviseksi luokiteltuja sanoja tulkitaan positiiviseksi sävyksi raportissa. Tutkimuksessa käytetyt sävykategoriat ovat positiivisuus, negatiivisuus sekä epävarmuus.

Tutkimuksen aiheena on talousraporttien ei-numeraalisen informaationsisällön merkitys sijoittajille. Tätä merkittävyyttä voidaan lähestyä informaation arvon sekä arvorelevanssin näkökulmista. Perinteisesti arvorelevanssitutkimus on koskenut laskentatoimen tuottamaa, usein numeraalista, tietoa talousraporteissa, jota on tarkasteltu osakekurssien vaihtelujen selittäjänä (ks. esim. Barth, Beaver & Landsman, 2001). Laajentamalla näkökantaa lukujen ylitse raportin koko informaationsisältöön vaihtelua voidaan selittää myös muilla raporttien elementeillä. Lähtökohtana on, että mikäli yrityksen julkaisema informaatio on arvorelevanttia, se aiheuttaa jonkinlaisen reaktion sijoittajissa, mikä näkyy osakkeen hinnan muutoksena (Scott, 2015). Laskentatoimen tuottamaa tietoa on tarkasteltu arvorelevanssin näkökulmasta jo pitkään (ks. esim. Ball & Brown, 1968). Tässä tutkimuksessa kuitenkin pyritään analysoimaan, onko yrityksen tavalla kuvata toimintaansa vaikutusta sijoittajien odotuksiin ja

arvioihin. Muiden kuin taloudellisten mittarien lisääminen laskentatoimen tulosten ohelle saattaa kasvattaa raportoinnin arvorelevanssia (Amir & Lev, 1996). Lisäksi aineettomaan pääomaan ja kasvumahdollisuuksiin liittyvä informaatio vaikuttaisi olevan yhä suuremmassa roolissa raportoinnin arvorelevanssin kannalta, samalla kun perinteisemmät mittarit, kuten liikevoitto, menettävät merkitystään osakekurssien vaihteluiden selittäjinä (Barth, Li & McClure, 2021). Sannallinen informaatio voidaan nähdä samanlaisena, kvantitatiivista informaatiota tukevana, tiedonlähteenä, jonka avulla yritysjohto voi täydentää ja selventää laskentatoimen tietoa. Julkaistun raportin sävyn voidaan olettaa olevan yrityksen keino tuoda oma näkemyksensä omasta toiminnastaan, mikä tuo lisää informaatiota sijoituspäätösten tekemiseen pelkkien laskentatoimen tuottamien lukujen vertailemisen ohelle. Huomionarvoista on kuitenkin tekstimuotoisen tiedon epävarmuus esimerkiksi tilintarkastettuihin laskelmiin verrattuna, sillä yritysjohto voi lausunnoillaan pyrkiä vaikuttamaan osakkeen arvon kehitykseen (ks. esim. Healy & Palepu, 2001), mikä voi heikentää tekstuaalisen informaation luotettavuutta.

Toisin kuin numeraalista talousdataa, yritysjohtoon kuvailevien lausuntojen sisältöä voi olla vaikea kvantifioida. Siksi yleisesti talousraportoinnin ja tekstianalyysin tutkimuksen haasteena on löytää tapa kuvata määrällisesti julkaistun raportin sävyä sen tekstisisällön perusteella. Aikaisemman tutkimuksen (ks. esim. Loughran & McDonald, 2011) tapaan tässä tutkimuksessa käytetään sanalista dokumentin sävyn määrittämiseen. Raporteista etsitään sanalistan sanoja, jotka on luokiteltu sävykategorioihin, kuten positiivisuus. Dokumentin sävy määritetään siinä esiintyvien sävykategorioihin kuuluvien sanojen lukumäärän mukaan, eli esimerkiksi kukin positiiviseksi luokiteltu sana kasvattaa raportin laskennallista positiivista sävyä. Käytännössä siis tarkastellaan, kuinka tiettyjen sanojen esiintyvyys raporteissa on yhteydessä markkinoiden reaktioon lyhyellä aikavälillä.

Yksi tutkimuksen tavoitteista on tutkia, onko yritysjohtoon arvioilla yrityksen tilasta merkitystä sijoittajien näkökulmasta, toisin sanoen tuovatko yrityksen julkaisemat näkymät lisäinformaatiota sijoittajille, ja voiko niitä hyödyntää sijoituspäätöksissä lyhyellä aikavälillä. Toisaalta asiaa voi ajatella myös sijoittajien näkökulmasta, jolloin voidaan pohtia, pystyvätkö sijoittajat hyödyntämään raporteissa olevan informaation tehokkaasti, eli onko sijoittajalla mahdollisuus keskimääräistä suurempiin lyhyen aikavälin tuottoihin raporttien sävyä analysoimalla. Toinen tutkimuksen tavoitteista on aiemmin pääasiassa amerikkalaisille yrityksille sovellettujen menetelmien laajentaminen suomalaiseen kontekstiin, jolloin tutkimus on luonteeltaan aikaisempaa tietoa vahvistava tai mahdollisesti kumoava.

Etenkin negatiivisen sävyn ja lyhyen aikavälin epänormaalien osaketuottojen välillä on havaittu yhteys, jossa negatiivisuuden kasvaessa epänormaalit tuotot vaikuttaisivat laskevan, mutta positiivisen sävyn yhteys sijoittajien reaktioon on nähty epävarmempana, osin sen vaikean määriteltävyyden takia, sillä myös negatiivisia uutisia saatetaan esittää positiivisin sanankääntein (Loughran & McDonald, 2016). Tämän lisäksi Cohen, Malloy ja Nguyen (2020) havaitsivat,

että lisäykset saman yrityksen raporteista toiseen ovat tyypillisemmin negatiivisia, mikä voi johtaa negatiivisen sävyn suurempaan merkittävyyteen. Sen vuoksi kiintoisaa on tutkia, mitkä toiminnan kuvauksessa esiintyvät sanat ovat olennaisia nimenomaisesti positiivisten sanojen yhteydessä ja siten aiheuttavat reaktion osakemarkkinoilla. Aihepiirit voivat koostua Lin (2010) tavoin esimerkiksi liikevaihtoon, kustannuksiin, investointeihin tai operatiiviseen toimintaan liitettävissä olevista sanoista.

Edellä esitetyistä ongelmista ja tavoitteista voidaan muodostaa seuraavat tutkimuskysymykset:

- Onko suomalaisten talousraporttien sävyllä ja niiden julkaisuhetkellä muodostuvilla lyhyen aikavälin epänormaaleilla osaketuotoilla yhteyttä?
- Mitkä sanat tai sanaryhmät positiivisten sanojen lähistöllä ovat yhteydessä poikkeavaan reaktioon osakemarkkinoilla?

Tutkimuskysymyksiin vastataan tarkastelemalla tapahtumatutkimuksen menetelmällä osavuositarkastusten julkaisuhetkellä syntyneitä lyhyen aikavälin epänormaaleja tuottoja, joita selitetään tekstianalyysin avulla muodostetuilla sävykategorioilla. Selittäminen tapahtuu monimuuttujaregressioanalyysin avulla, jossa selittävinä muuttujina ovat sävykategoriat sekä kontrollimuuttujat. Lähisanojen määrittämiseksi kerätään lista sanoista, jotka positiivisen sanan yhteydessä esiintyessään aiheuttavat implikaatioidensa tai tulkintojensa seurauksena tilastollisesti merkitsevän eron keskimääräisissä epänormaaleissa tuotoissa. Tätä listaa analysoidaan laadullisesti teemoittelun menetelmällä yhtenäisyyksien löytämiseksi.

Tämän tutkielman kannalta keskeisiä käsitteitä ovat tekstianalyysi ja dokumentin sävy, jotka määritellään seuraavaksi. Tekstianalyysi tai sisällön erittely on tapa laadullisen aineiston kuvailemiseksi kvantitatiivisin keinoin. Tyypillisesti se pitää sisällään sanojen esiintymistiheyksien laskemista dokumenteista. (Tuomi & Sarajarvi, 2018.) Tällaisissa menetelmissä sanojen järjestyksen tai niiden keskinäisten suhteiden merkitystä ei oteta huomioon, vaan olennaisena pidetään ainoastaan sitä, että jokin termi tai sana on esiintynyt aineistossa. Sisällön määrällistäminen on mahdollista tehdä tietokoneavusteisesti, mikä helpottaa suurten aineistojen käsittelyä. (Loughran & McDonald, 2016.)

Dokumentin sävy määritellään tässä tutkimuksessa sisällön erittelyn menetelmällä. Tässä menetelmässä dokumentin sisältöä tarkastellaan siinä olevien sanojen frekvenssien avulla niiden keskinäisistä yhteyksistä välittämättä (Loughran & McDonald, 2016). Sävy määräytyy positiiviseksi, negatiiviseksi tai epävarmuutta kuvaaviksi luokiteltujen sanojen osuuksien mukaan. Mitä enemmän dokumentin sanoista on luokiteltu positiiviseksi, sen korkeamman arvon dokumentti saa positiiviseksi sävyluokitukseen. Sama prosessi toistuu muillakin tarkasteltavilla kategorioilla.

Tämän johdantoluvun jälkeen tutkielma etenee seuraavasti. Ensin esitellään tutkielman kannalta olennainen teoreettinen viitekehys, jossa aluksi tarkastellaan talousraportointia, osavuositarkastuksia ja niiden vaikutuksia, minkä jälkeen arvioidaan raportoinnissa esiintyvää sanallista informaatiota. Teoriaosuus päättyy laskentatoimen alalla tehdyn tekstianalyysiin perustuvan tutkimuksen

tarkasteluun. Tämän jälkeen käydään läpi aineiston sisältö ja sen keräämisen sekä prosessoinnin vaiheet, ennen kuin paneudutaan tutkimuksessa käytettäviin menetelmiin ja aineiston analysointiin. Lopuksi vedetään yhteen johtopäätöksiä sekä pohditaan tutkimuksen rajoitteita ja jatkotutkimusaiheita.

2 TALOUSRAPORTOINTI

2.1 Yleistä

Talousraportointi voidaan nähdä ratkaisuna kahteen ongelmaan, jotka koskevat sijoittajan ja yrityksen välistä sopimista. Ensimmäinen niistä on informaatioasymmetria, jossa ongelmana on yrityksen ja sijoittajan hallussa pitämän informaation ero. Jos sijoittajalla on huomattavasti vähemmän tietoa yrityksen tilasta kuin yrityksen henkilöstöllä, toisin sanoen informaatioasymmetria on suuri, sijoittaja ei pysty tekemään arviota yrityksestä ja siten rationaalista sijoituspäätöstä. (Healy & Palepu, 2001.) Toisaalta informaatioasymmetrian avulla yritysjohtolla on mahdollisuus tehdä sijoittajan kannalta potentiaalisesti haitallisia ratkaisuja, joista sijoittaja ei tule tietoiseksi. Talousraportointi pienentää informaatioasymmetriaa uskottavalla ja oikea-aikaisella informaation julkaisulla samalla vähentäen riskiä sijoittajien arvioinneissa. (Scott, 2015.) Toinen ongelma, johon talousraportointi puuttuu, on agenttiongelmia. Sen mukaan sijoittaja antaa rahamääräisen panoksena yritykselle, mutta ei kuitenkaan osallistu aktiivisesti yrityksen toimintaan, jolloin yritysjohtolle saattaa muodostua mahdollisuus tehdä omaa etua tavoittelevia päätöksiä sijoittajan edun kustannuksella. Talousraportoinnin avulla sijoittajan on mahdollista seurata yrityksen toimia, vaikka ei yrityksen toimintaan aktiivisesti puuttuisikaan, mikä vähentää agenttiongelmien perustuvaa epävarmuutta. (Healy & Palepu, 2001). Nämä periaatteet näkyvät myös suomalaisessa lainsäädännössä. Suomalaista listattua yhtiötä sitoo arvopaperimarkkinalain mukainen säännöllinen tiedonantovelvollisuus, eli pörssiyhtiöltä vaaditaan säännöllistä raportointia esimerkiksi tilinpäätöksen tai osavuosikatsauksen muodossa. Tämän seurauksena pääomamarkkinoiden tulisi tehostua. (Parkkonen & Knuts 2014, 144.)

Tärkeimpänä talousraportoinnin laatua ohjaavana periaatteena on pidetty tiedon hyödyllisyyttä päätöksenteon kannalta, jonka mukaan jaetun informaation ei tarvitse olla teoreettisen täsmällisesti oikeaa, mutta sen tulee olla tarpeeksi todenmukaista päätöksentekoa varten (Scott, 2015). Tämä näkyy myös suomalaisissa tilinpäätöskäytänteissä esimerkiksi olennaisuuden periaatteena,

jonka perusteella voidaan määrittää tilinpäätöksessä julkaistavien tietojen määrä ja laatu (Haaramo, Palmuaro & Peill, 2018). Informaatiota voidaan pitää olennaisena, jos sen ilmoittamatta jättämisen seurauksena sijoittajan rationaalinen päätöksenteko estyisi. Olennaisuuden lisäksi raportoidun tiedon tulisi olla todenmukaista ja riittävää, minkä lisäksi sen julkaisun on oltava johdonmukaista tilikausien välillä sekä vastaanottajien kannalta tasapuolista. (Parkkonen & Knuts 2014, 145.)

2.1.1 Talousraportoinnin merkitys sijoittajan kannalta

Yritysten julkaiseman informaation merkitystä osakemarkkinoilla voidaan tarkastella tehokkaiden markkinoiden hypoteesin avulla. Faman (1970) mukaan tehokkailla markkinoilla on kolme muotoa: heikko, keskivahva sekä vahva. Tyypillisesti markkinoiden oletetaan olevan keskivahvasti tehokkaat, jolloin osakkeiden hinnat perustuvat kaikkeen julkisesti saatavilla olevaan informaatioon, ja uuden relevantin informaation tullessa markkinoille sijoittajien reaktio on välitön (Scott, 2015). Heikon tehokkuuden markkinoilla hinnat perustuvat historialliseen kehitykseen, kun taas vahvoilla markkinoilla hinnoittelussa on otettu huomioon kaikki informaatio, myös sisäpiirin hallussa oleva yksityinen tieto (Fama, 1970). Tehokkaiden markkinoiden hypoteesi perustuu oletukseen taitavasta ja rationaalisesta sijoittajasta, joka onnistuu purkamaan ja arvioimaan yrityksen raportoinnin sisällöstä yrityksen tulevaa suorituskykyä kuvaavaa informaatiota. Toisen näkökulman sijoittajien ominaisuuksiin tuo toiminnallisen fiksaation hypoteesi, jonka mukaan sijoittajat eivät kykene erottamaan raportoinnista tarvittavaa informaatiota, vaan ovat alttiita esimerkiksi yrityksen valitsemien laskentatapojen eroavaisuuksille. (Hand, 1990). Handin (1990) hypoteesi yhdistää nämä kaksi näkökulmaa, sillä sen mukaan osakkeiden hinnat perustuvat vain osin taitavien sijoittajien valintoihin, jolloin taitamattomien sijoittajien arviot vaikuttavat omalta osaltaan hinnanmuodostukseen.

Olennaista talousraportoinnin ja osakkeiden hinnanmuodostuksen kannalta on myös sijoittajien aktiivinen informaation kerääminen. Grossmanin ja Stiglitzin (1976) teorian mukaan osakkeiden hinnat eivät vastaa täysin saatavilla olevaa informaatiota, sillä kaikki sijoittajat eivät kerää informaatiota yhtä kattavasti tai perustavat päätöksensä muuhun kuin yrityksestä saatavilla olevaan tietoon. Näin ollen informaation keräämisen kannustin sijoittajan kannalta on tarkentuneen hinnan tuloksena syntynyt tuotto (Grossman & Stiglitz, 1976). Talousraportointi on lähtökohtaisesti julkista informaatiota, mutta taitavat sijoittajat saattavat pystyä tekniikoillaan jalostamaan raportoinnissa olevista tiedoista yksityistä tietoa (Kim & Verrecchia, 1994). Holdenin ja Subrahmanyamin (1992) mukaan kuitenkin myös yksityinen informaatio sisäistyy osakkeiden hintoihin nopeasti, kunhan saman informaation haltijoita on enemmän kuin yksi. Epäilyksen alaista talousraportoinnin seurauksissa osakemarkkinoilla on ollut se, kuinka hyvin sijoittajat pystyvät hyödyntämään kaikkea raportoinnissa esiintyvää informaatiota. Hirschleiferin ja Teohin (2003) mukaan jokaisella sijoittajalla on jokin todennäköisyys sisäistää raportoinnissa julkaistu informaatio täydellisesti, jolloin osia julkisesta informaatiosta voi jäädä huomiomatta sijoit-

tuspäätöksissä. Merkittävä ja helposti ymmärrettävässä muodossa julkaistu informaatio sisäistetään helpommin, mutta kaikki yksityiskohdat, varsinkin epäsuorasti ilmi tulevat, eivät välttämättä saa sijoittajien huomiota (Hirschleifer & Teoh, 2003). Esimerkkinä tästä ovat tietyt sanallisessa muodossa esitetyt arviot, jotka saattavat siirtyä osakkeiden hintoihin viiveellä (Cohen ym., 2020).

Kaikki raportoinnissa julkaistu informaatio ei välttämättä ole hyödyllistä sijoittajien päätöksenteon kannalta. Arvorelevanssitutkimuksen tavoitteena on selvittää laskentatoimen tuottaman tiedon merkitystä sijoittajien kannalta, toisin sanoen pystyvätkö erilaiset laskentatoimen määreet tai käsitteet selittämään osakekurssien vaihtelua (Barth ym., 2001). Talousraportoinnissa esitettyjen tietojen on todettu siirtyvän osakkeiden hintoihin ja olevan siten arvorelevantteja (ks. esim. Ball & Brown, 1968). Kuitenkin Lev ja Zarovin (1999) havaitsivat muutamien tärkeiden taloudellisten mittareiden, kuten liikevoiton, kassavirtojen ja taseen loppusumman, hyödyllisyyden pienentyneen sijoittajien päätöksenteon näkökulmasta ajan saatossa. Myös Francis ja Schipper (1999) päätyivät samaan tulokseen liikevoiton osalta, mutta taseen varat ja velat puolestaan vaikuttivat olevan yhä relevantteja osakkeiden arvomuodostuksen kannalta. Yhtenä syynä talousraportoinnin arvorelevanssin vähenemiseen on pidetty aineetoman pääoman merkityksen kasvamista, jolloin perinteisemmät tunnusluvut ovat suhteessa menettäneet merkitystään eivätkä enää kuvaa yrityksen tilaa tarpeeksi tarkasti (Lev & Zarovin, 1999; Lev, 2019). Onkin todettu, että kokonaisuudessaan talousraporttien arvorelevanssi ei ole heikentynyt, sillä erilaiset kasvua, aineetonta pääomaa tai muita vaihtoehtoisia suoritusmittareita kuvaavat erät ovat nousseet olennaisemmiksi sijoittajien päätöksenteon kannalta, ja talousraportoinnin sekä osakkeen hinnanmuodostuksen välinen yhteys on muuttunut tämän seurauksena monipuolisemmaksi (Barth ym., 2021).

2.1.2 Osavuosikatsaus

Eräs talousraportoinnin muodoista on säännöllisesti, yleensä neljännesvuositain, tapahtuva osavuosiraportointi. Suomalaisilta pörssiyrityiltä edellytetään tilinpäätöksen sekä puolivuotiskatsauksen julkaisemista, mutta ensimmäiseltä ja kolmannelta neljännekseltä tehtävän raportin laatiminen on vapaaehtoista. Osavuosikatsauksen rooli on aikaisempia tilinpäätöksiä päivittävä, joten se on tarkoitettu uuden informaation julkaisemiseen edellisten tulosten toistamisen asemesta. (Haaramo ym., 2018.) Kansainvälisistä tilinpäätösstandardeista IAS 34 koskee osavuosikatsauksia, vaikkakin se kohdistuu pääosin raporttien taulukko-osuuteen, kun taas osavuosikatsausten selostavaan osaan standardi ei ota juurikaan kantaa. Taulukko-osan sisältö perustuu kunkin jakson olennaisiin tapahtumiin, tosin arvioihin perustuvia laskelmia siedetään hiukan koko vuotta koskevia raportteja enemmän. Suomalaisten yritysten osalta osavuosikatsauksen selostusosan sisältöä määrittelee arvopaperimarkkinalaki, jonka mukaan kyseisessä osassa on annettava yleiskuvaus tarkastelujakson taloudellisista tapahtumista ja kehityksestä. Lisäksi on selvitettävä mahdollisia lähitulevaisuuden riskitekijöitä, mutta varsinaisia toimintaa ennakoivia lausuntoja ei tarvitse antaa. (Parkkonen & Knuts 2014, 191–197.) Osavuositaisen raportoinnin tavoit-

teina voidaan pitää ennusteiden parantamista, toiminnan muutosten paikallistamista, johdon arvioinnin kehittämistä sekä vuosittaisen tilinpäätöksen tukemista (McEwen & Schwartz, 1992).

Seuraavaksi tarkastellaan osavuosikatsauksia niiden merkittävyyden ja sopivan raportointivälin kannalta. Sopivaan raportointitiheyteen liittyy olennaisesti ajatus sosiaalisesti optimaalisesta informaatiomäärästä, jossa tiedon julkaiseminen tai sen hankkiminen ei muutu liian kalliiksi yritysten tai sijoittajan kannalta. Toisaalta myöskään liian harvaan tapahtuva raportointi ei ole sosiaalisesti optimaalista, mikäli se aiheuttaa lieveilmiöitä, kuten informaatioasymmetriaa tai liiasta arvioinnista seuraavia riskejä. (Scott, 2015.) Osavuosikatsausten merkittävyyttä puolestaan on tutkittu esimerkiksi osakemarkkinoiden reaktioiden tai analyytikkojen seurannan valossa (Kajüter, Lessenich, Nienhaus & van Gemmern, 2021).

Analyytikkojen toimintaan osavuosikatsausten julkistaminen saattaa tuoda lisäkustannuksia, mikä voi johtaa joidenkin yritysten osalta pienempään seurantaan ja sitä kautta heikompaan informaatioympäristöön (Yee, 2004). Kajüter ym. (2021) kritisoivat kuitenkin Yeen (2004) tuloksia, sillä ne eivät ota huomioon analyytikkojen informaationkeräyksen tehostumista tiheään raportoinnin myötä. Lisäksi tiheällä raportoinnilla saattaa olla näkyvyyttä tai julkisuutta tuova vaikutus, joka voisi lisätä analyytikkojen seurantaan yritystä kohtaan (Kajüter ym., 2021). Osavuosikatsauksen tai muun vastaavan raportin julkistaminen saattaakin olla analyytikolle tai muulle edistyneelle sijoittajalle tilaisuus hyödyntää omia yksityisiä tietojaan keskivertosijoittajaa tehokkaammin (Kim & Verrechia, 1994). Tämä taas voi johtaa analyytikkojen ennusteiden parantumiseen (ks. esim. Brown & Rozeff, 1979).

Osavuosikatsaukset vaikuttaisivat olevan olennaisia tietolähteitä sijoittajien kannalta osakemarkkinoilla tapahtuvien reaktioiden perusteella. Etenkin raporttien tultua sähköisesti ladattaviksi sijoittajien on todettu reagoivan pikaisesti niiden tarjoamaan informaationvälitykseen, mikä näkyy osakkeiden arvon muutoksina raportin julkaisuajankohtana (Griffin, 2003). Griffinin (2003) amerikkalaisia yhtiöitä koskeva tulos on toistettu myös eurooppalaisilla yrityksillä (ks. esim. Alves & Dos Santos, 2008). Osavuosikatsausten aiheuttama reaktio osakemarkkinoilla ei kuitenkaan ole yhtä merkittävä, jos niitä edeltää jokin muu vastaavanlaisen informaation julkistus (Li & Ramesh, 2009). Yritykset saattavatkin julkaista informaatiota muussa muodossa, kuten pörssitiedotteina, kuin kokonaisuutena osavuosikatsauksena ja näin korvata tarpeen jaksottaiselle neljännesraportoinnille (Schleicher & Walker, 2015).

Sijoittajien reagoinnin olemassaolon lisäksi osavuosikatsauksia on tarkasteltu reaktioiden nopeuden kannalta. Mikäli reaktio on nopea, eli uusi informaatio sisällytetään osakkeiden hintaan täsmällisesti ajan suhteen, markkinoita voidaan pitää tehokkaina (Kajüter ym., 2021). Osavuosikatsausten julkaiseminen neljännesvuosittain ei kuitenkaan näyttäisi parantavan osakemarkkinoiden tehokkuutta ainakaan hinnanmuodostuksen nopeuden kannalta verrattuna puolivuositaiseen julkaisuväliin (Butler, Kraft & Weiss, 2007). Tsao, Lu & Keung (2018) saivat kuitenkin toisenlaisia tuloksia, joiden mukaan pienempi

raportointiväli tehostaisi hinnanmuodostusta. Heidän tutkimuksensa on kuitenkin rajoittunut Taiwaniin, mikä taloudellisten erityispiirteittensä vuoksi saattaa tehdä tuloksista vaikeammin yleistettäviä (Kajüter ym., 2021). Tiheämmin raportoivilla yrityksillä julkaisuajankohdan osakkeen hinnan vaihtelu on todettu olevan suurempaa, mikä voisi viitata heikompaan tehokkuuteen hinnanmuodostuksen kannalta (Mensah & Werner, 2008). Suureen vaihteluun syynä voi olla sijoittajien ylireagointi osavuositarkastuksissa esitettyihin arvioihin, jotka saattavat olla esimerkiksi tilinpäätöksiä epävarmempia ja usein myös tilintarkastamattomia (Kajüter ym., 2021). Toisaalta, mikäli tiheämpi julkaisu tahti on seurausta yrityksen oma-aloitteisuudesta, eikä esimerkiksi lainsäädännöllisestä pakosta, informaatio näyttäisi sisältyvän osakkeen hinnoitteluun tehokkaammin (Butler ym., 2007). Hinnanmuodostusta saattaa hidastaa raportoinnin monimutkaisuus, sillä vaikeaselkoisempien osavuositarkastusten lyhyellä aikavälillä aiheuttama reaktio on helpommin ymmärrettäviä maltillisempi. Tämä viittaisi hitaampaan tiedon sisäistämiseen ja analysointiin monimutkaisien tai vaikeasti luettavien raporttien osalta. Informaatioympäristön laatu voi vaikuttaa raportoinnin luettavuuden vaikutuksiin, sillä jos julkaiseva yritys on analyytikoiden seurannassa, monimutkaisuudesta koituva hinnanmuodostuksen hidastuminen ei ole yhtä voimakasta. (Lee, 2012).

Vaikka tiheämpi raportointiväli neljänneskatsausten muodossa ei parantaisikaan hinnanmuodostuksen tehokkuutta, sen käytöstä on löydetty muita mahdollisia hyötyjä. Taajan raportoinnin on empiirisesti todettu vähentävän informaatioasymmetriaa (Fu, Kraft & Zhang, 2012), vaikkakin teoreettisesti asiasta on esiintynyt vaihtelevia näkökantoja. Diamondin (1985) mukaan asymmetrian pitäisi pienentyä tiheän raportoinnin seurauksena, kun sijoittajille tarjotaan useammin mahdollisuus julkiseen informaatioon tasavertaisesti. Informaatioasymmetria saattaa kuitenkin tästä huolimatta kasvaa, mikäli jotkin tahot markkinoilla pystyvät prosessoimaan ja analysoimaan julkisesta informaatiosta yksityistä tietoa muita tehokkaammin, ja usein tapahtuva raportointi antaa enemmän mahdollisuuksia tähän (Kim & Verrechia, 1994). Fu ym. (2012) laskevat myös pääoman kustannusten alenemisen suuremman raportointitiheyden hyödyksi. Samaan lopputulemaan päätyivät myös Francis, Nanda ja Olsson (2008), tosin heidän mukaansa pääoman kustannukset laskevat vain, jos yrityksen julkaisema informaatio on laadukasta, eli yrityksen talousraportointijärjestelmä pystyy pitämään julkaisemansa informaation riittävän tarkasti oikeaa kokonaiskuvaa vastaavana.

Yhtenä osavuositarkastuksien ongelmista voidaan pitää niiden keskinäistä samankaltaisuutta, jolloin uutta relevanttia informaatiota ei jokaisella raportointikerralla tule riittävästi julkaistuksi. Lightstone, Young & McFadden (2012) havaitsivat, että kolmen ensimmäisen neljänneksen raporteissa liikevoiton vaihtelu on pienempää kuin tilikauden lopussa julkaistavassa katsauksessa, mikä voisi viitata siihen, että kesken tilikauden kootut raportit eivät ole tarpeeksi tarkoin muodostettuja, ja tilikaudella tapahtuvia muutoksia tarkastellaan tosiasiasa vasta tilikauden lopussa. Lisäksi tilikauden aikana tapahtuvan negatiivisen informaation julkaisemista saatetaan lykätä tilikauden loppuun, jolloin tilinpä-

töksen tilintarkastus tekee julkaisemisesta välttämätöntä (Pastena & Ronen, 1979). Toisaalta, mikäli julkaistava informaatio on monimutkaista ja sen valmistelu vie aikaa, voi raportin julkaiseminen viivästyä, jolloin julkaistaessa informaatio saattaa olla vanhentunutta (Kajüter ym., 2021). Näyttäisikin siltä, että raportin julkaisun lykkääntyessä reaktio osakemarkkinoilla on ajankohtaista raportointia maltillisempi (Beaver, McNichols & Wang, 2018). Kuitenkin esimerkiksi Helsingin pörssin sääntöjen mukaan yhtiöiden on päätettävä ennalta julkaisuaikataulunsa (Parkkonen & Knuts 175, 2014), joten suomalaisten yritysten tapauksissa kyse saattaa olla ennemminkin julkaisun siirtämisestä seuraavalle raportointikerralle, kuin raportin julkaisupäivämäärän lykkäämisestä. Yhtenä ratkaisuna osavuosisikatsausten relevanssiongelmiin voisi olla niiden tilintarkastaminen, joka ei kaikissa maissa ole välttämätöntä (Lightstone ym., 2012). Tämän seurauksena osavuosisiraportoinnin merkittävyys sijoittajien näkökulmasta saattaisi kasvaa sekä raporttien todenmukaisuuden parantumisen että tilintarkastamisen luoman positiivisen signaalin avulla (Kajüter, Klassmann & Nienhaus, 2016).

Osavuosisikatsauksien laatimisen hyödyt näyttäisivät olevan kokonaisuudessaan niiden haittoja suurempia. Niiden tarjoama sisältö otetaan vastaan osakemarkkinoilla sekä analyytikkojen toimesta, minkä lisäksi niiden avulla saataan voida parantaa pääomamarkkinoiden tehokkuutta esimerkiksi pääoman kustannuksia alentamalla, vaikka joitakin tehokkuutta alentavia ilmiöitä, kuten informaatioasymmetrian kasvu, on yhdistetty osavuosisiraportointiin. Huolta aiheuttaa myös raporttien relevanssi ja todenmukaisuus esimerkiksi vuosittaiseen tilinpäätökseen nähden, mihin ratkaisuna voisi olla ulkoisen tilintarkastuksen lisääminen mukaan osavuosisittaiseen raportointiprosessiin.

2.2 Sanallinen informaatio raportoinnissa

Talousraportit ovat informaatiolähteinä monipuolisia, sillä ne sisältävät tietoa sekä numeraalisessa että sanallisessa muodossa. Pelkkien taloudellisten lukujen julkaiseminen ei kuitenkaan välttämättä ole riittävää vaadittavan taloudellisen kuvan antamiseksi, varsinkaan nopean kasvun aloilla (Amir & Lev, 1996). Talousjohdon oma analyysi yrityksen toiminnasta saattaakin olla taloudellista dataa olennaisempi informaatiolähde, sillä Rogersin ja Grantin (1997) mukaan suurin osa ulkopuolisten analyytikkojen julkaisujen viittauksista on peräisin tilinpäätöksiensä osasta, jossa talousjohto kertoo omia näkemyksiään. Siksi talousraporteissa olevaa toimintaa kuvailevaa ja analysoivaa osaa voidaan pitää yrityksen kannalta mahdollisuutena kattavampaan informaation julkaisemiseen ja siten raportointinsa arvorelevanssin kasvattamiseen (Scott, 2015). Esimiehillä on mahdollisuus tuoda lausunnoillaan esiin muutoin mahdollisesti vähälle huomiolle jäävää informaatiota (Merkl-Davies & Brennan, 2007). Lisäksi yritysjohdolle voi sanallisesti yhdistää raportoituja tuloksia sisäisiin tai ulkoisiin syihin, minkä tavoitteena on sijoittajan päätöksenteon tukeminen (Baginski, Hassell & Kimbrough, 2004). Toiminnan tarkka sanallinen kuvaaminen voidaan nähdä

ylimääräisenä tai vapaaehtoisena julkistuksena, joka luo osaltaan kustannuksia (Scott, 2015). Sen vuoksi seuraavaksi tarkastellaan syitä, minkä vuoksi yrityksen kannattaa tuoda enemmän tietoa julki kuin esimerkiksi lainsäädäntö edellyttää. Toisaalta samaan aikaan tulee ottaa huomioon mahdollisuus vilpilliseen tai opportunistiseen toimintaan esimerkiksi tulostavoitteiden saavuttamiseksi (ks. esim. Healy & Palepu, 2001), minkä vuoksi käydään läpi myös sanallisen raportoinnin luotettavuuteen vaikuttavia tekijöitä.

Teoreettisesta näkökulmasta osakkeensa arvoa maksimoivan yrityksen kannattaa julkaista kaikki relevantti informaatio sen toiminnasta, kunhan muuttamat ehdot täyttyvät. Ensinnäkin yritys tietää, miten sijoittajat reagoivat informaation julkaisuun ja voi olettaa, että reaktio on samanlainen kaikilla sijoittajilla. Toisekseen sijoittajilla tulee olla tieto siitä, että yrityksellä on yksityistä ja vielä julkaisematonta informaatiota. Lisäksi yrityksellä tulee olla tapa julkaista informaatiota uskottavasti, mutta samalla kustannuksettomasti. Käytännössä yritykset eivät kuitenkaan julkaise kaikkea yksityistä informaatiotaan, mikä viittaa jonkin ehdon täyttymättömyyteen näissä tapauksissa. (Beyer, Cohen, Lys & Walter, 2010.) Mikäli nämä ehdot täyttyvät jonkin tiedonkappaleen osalta ja se julkistetaan, ylimääräistä raportointia voidaan pitää lisätietoa tuovana (Merkl-Davies & Brennan, 2007). Yleisellä tasolla informaation julkistaminen pienentää arvioimisesta koituvaa riskiä sijoittajien kannalta, joten ylimääräinen julkaiseminen esimerkiksi sanallisen informaation muodossa parantaa yrityksen arvostamisen tarkkuutta (Scott 2015, 141). Toisaalta sijoittajat saattavat yhdistää julkaisematta jättämisen tai raportoinnin puutteellisuuden huonojen uutisten piilotteluksi, mikä johtaa alentuneeseen arvostukseen (Milgrom, 1981). Sen vuoksi myös huonojen uutisten raportointi voi olla yrityksen osalta kannattavaa lainsäädännöllisten vaatimusten lisäksi sen tuottaman luotettavuuden ja oikeudellisuuden signaalin takia (Yuthas, Rogers & Drillard, 2002). Huonoja uutisia saatetaan julkaista ennakkoon, jotta suurilta negatiivisilta yllätyksiltä vältyttäisiin varsinaisena raportointiajankohtana (Skinner, 1994).

Raportoivalla yrityksellä on teoreettisen informaation jakamisen maksimoinnin tavoitteen lisäksi käytännössä todettuja syitä sanalliseen tiedonjulkaisuun. Tauringana ja Mangena (2006) jakavat sanallisen informaation raportoinnissa sen mukaan, liittyykö tieto tilinpäätöksen lukuihin vai johonkin muuhun seikkaan, esimerkiksi yleisempiin tulevaisuuden näkymiin. Lukuja täydentäviä tietoja antamalla yritysjohto voi selittää tapahtumia tulosten takana esimerkiksi edeltävään tarkastelujaksoon verrattuna (Tauringana & Mangena, 2006). Laskentatoimen tuottama numeraalinen tieto on luonteeltaan yleistävää ja objektiivista, mutta sanallinen informaatio mahdollistaa raportoinnin yksilöllistämisen, sillä sen avulla yritysjohto pystyy perustelemaan tekemiään päätöksiä ja selittämään toimintansa tuloksia (Aerts, 1994). Hyvän tuloksen tehneillä on mahdollista erottautua muista yrityksistä tuomalla esiin lisätietoa onnistumisen taustoista (Skinner, 1994). Toisaalta huonojen tulosten yhteydessä saatetaan tuoda sanallisesti esiin muita yrityksen kannalta olennaisia asioita, kuten tutkimuksen ja kehityksen saavutuksia (Merkley, 2014).

Jotta julkaistu informaatio on sijoittajien päätöksenteon kannalta hyödyllistä, sen täytyy olla tarpeeksi luotettavaa. Sanallisen sisällön luotettavuutta talousraportoinnissa voidaan tarkastella kahdesta vastakkaisesta näkökulmasta, joita ovat sen käyttäminen lisäinformaation tuottamiseen tai vaihtoehtoisesti sen avulla sijoittajan päätöksenteon häiritseminen (Merkl-Davies & Brennan, 2007). Jälkimmäisessä tapauksessa puhutaan vaikutelmien hallinnoimisesta, jossa yritysjohto käyttää talousraportointia sidosryhmiensä käsitysten ja päätösten manipuloimiseen (Yuthas ym., 2002). Nämä kaksi koulukuntaa perustuvat erilaisille oletuksille markkinoiden tehokkuudesta. Mikäli sijoittajat pystyvät näkemään kuvailevan informaation puolueellisuuden läpi, jolloin tämä informaatio on päätöksenteon kannalta hyödyllistä riippumatta sen sisältämästä keinottelusta, markkinoita voidaan pitää vahvasti tai keskivahvasti tehokkaina (Merkl-Davies & Brennan, 2007). Näin ollen yritysjohdolla ei olisi syytä raportoinnin sanalliseen manipulointiin, sillä sen sijaan, että se mahdollistaisi osakkeen arvoon vaikuttamisen, manipulointi ainoastaan heikentäisi yrityksen luotettavuutta ja arvostusta (Baginski, Hassell & Hillison, 2000). Toisen näkökannan mukaan yrityksellä on mahdollisuus muokata sijoittajien käsityksiä vaikutelmia hallinnoimalla, jolloin markkinat ovat muodoltaan heikot (Merkl-Davies & Brennan, 2007).

Sanallisen informaation hyödyllisyyden näkökulmaa on tuettu sijoittajien reaktioihin perustuen. Tällöin ajatuksena on, että raportointia ohjaa sanallisen informaation arvorelevanssi, ja sijoittajat eivät ole alttiita yrityksen keinottelulle raportoinnissa vaan pystyvät arvioimaan lausuntojen luotettavuutta rationaalisesti (Merkl-Davies & Brennan, 2007). Francis, Schipper ja Vincent (2002) havaitsivat, että osakemarkkinoiden reaktio raporttien julkaisun yhteydessä selittyy osin raporttien sisältämän lisäinformaation, johon laskettiin myös sanalliset arviot, avulla. Baginskin ym. (2004) mukaan yritysjohton antamat sanalliset selvennykset vaikuttavat sijoittajien reaktioihin, etenkin jos ne perustuvat ulkoisiin tekijöihin, kuten yleiseen taloustilanteeseen. Sijoittajien reaktio kuvailevaan informaatioon vaikuttaisi riippuvan raportoinnin muusta sisällöstä, sillä Hutton, Miller ja Skinner (2003) totesivat, että huonojen julkaistujen ennusteiden yhteydessä sanallinen lisäinformaatio ei vaikuttanut sijoittajien reaktioon, kun taas positiivisten ennusteiden yhteydessä reaktio on lähes olematon, ellei niiden ohella ole annettu todennettavissa olevia arvioita tulevaisuuden kehityksestä. Tämä viittaisi siihen, että sijoittajat uskovat helpommin yrityksen raporttoimia huonoja ennusteita, kun taas hyvät ennusteet vaativat jonkinlaista vahvistavaa informaatiota, jotta niitä pidetään uskottavina osakemarkkinoilla (Hutton ym., 2003). Edellä kuvattua tutkimushaaraa voidaan pitää samankaltaisten oletuksiensa vuoksi edeltäjänä tekstianalyysiin perustuville tutkimuksille, joita tarkastellaan tarkemmin luvussa 3.

Vaikutelmien hallinnoimista kuvaava tutkimus ottaa kriittisemmän näkökannan sanalliseen informaatioon raportoinnissa. Se nojautuu taloustieteiden käyttäytymistutkimukseen, jonka mukaan sijoittajat eivät pysty aina täysin rationaalisiiin päätöksiin, vaan päätöksentekoa saattavat ohjata erilaiset harhat ja epävarmuudet, minkä seurauksena sijoittajat ovat alttiita yrityksen raportoin-

nissa esiintyvälle vaikutelmien hallinnoimiselle. Vaikutelmien hallinnoimisen strategioita on kahta tyyppiä. Hämäämisellä tai piilottelulla voidaan korostaa positiivisia tuloksia ja vähentää negatiivisen informaation merkitystä, kun taas syitä tuloksiin liittämällä saatetaan korostaa yrityksen sisäisiä toimia onnistuneiden tulosten takana ja yhdistää heikot tulokset ulkoisiin seikkoihin. (Merkl-Davies & Brennan, 2007.)

Eräs raportoinnissa esiintyvä piirre, jolla sijoittajia saatetaan hämätä, on informaation esittämisen manipulointi. Esimerkkinä tästä on Schrandin ja Waltherin (2000) huomio, jonka mukaan yritykset saattavat esittää raportoidun liikevoiton vertailutuloksen jakamalla sen osiin, jolloin nykyinen tulos näyttää vahvemmalta vertailuarvoonsa nähden. Lisäksi Schrand ja Walther (2000) toteavat, että sijoittajat näyttäisivät käyttävän tätä muokattua vertailutulosta arvioinneissaan, mikä viittaisi siihen, että yrityksen onnistuu tulosten esitystapaa muokkaamalla vaikuttaa sijoittajien päätöksentekoon. Yritysjohdon on mahdollista vaikuttaa sijoittajien näkemyksiin myös informaation järjestystä muuttamalla, sillä raportin alkupuolella annettu tieto saattaa vaikuttaa sijoittajien näkemyksiin raportin muissa osissa sijaitsevia tietoja enemmän (Baird & Zelin, 2000). Lisäksi Boudt ja Thewissen (2019) havaitsivat, että positiivisia uutisia mainitaan eniten raporteissa julkaistavien toimitusjohtajan kommenttien alku- ja loppuosissa, kun taas negatiiviset uutiset painottuvat näiden tekstien keskivaiheille, mikä yhdistettynä käsitykseen tekstien alkujen ja loppujen tärkeämmästä merkityksestä lukijan havainnoinnin kannalta viittaisi yritysjohdon tietoiseen informaation jäsentämiseen. Kaikkiin sijoittajiin informaation esittämisestä aiheutuva harha ei vaikuta yhtä paljon, sillä kokemattomimmat sijoittajat ovat alttiimpia esitystavasta johtuvaan väärään päättelyyn ammattimaisiin sijoittajiin verrattuna (Elliot, 2006).

Toinen keino, jolla yritykset saattavat pyrkiä hämäämään sijoittajien näkemyksiä, on valikoiva raportointi, jossa yritysjohdolla on mahdollisuus päättää joidenkin tietojen sisällyttämisestä tai poisjättämisestä raportoinnissaan (Merkl-Davies & Brennan, 2007). Leungin, Parkerin ja Curtisin (2015) mukaan yritykset, jotka raportoivat sanallisesti vähiten kattavasti, ja siten jättäen informaatiota mahdollisesti raportoimatta, ovat keskimäärin heikommassa taloudellisessa tilanteessa kuin laajemmin toiminnastaan kertovat. Eniten eroja laajasti ja suppeasti raportoivien välillä on yrityksen strategiaa, tavoitteiden saavuttamista ja toimintojen yleiskatsausta koskevilla raporttien osilla, mikä viittaisi siihen, että suppeasti raportoivat yritykset pyrkivät vähentämään sijoittajien saamaa informaatiota koskien yrityksen hallinnollista toimintaa (Leung ym., 2015). Raportoinnin suppeus ei kuitenkaan välttämättä viittaa strategiseen vaikutelmien hallinnointiin, sillä esimerkiksi omistusrakenteella on todettu olevan vaikutusta raportoinnin laajuuteen (Hadro, Klimczak & Pauka, 2017).

Julkaistavan informaation valikoimisen lisäksi sijoittajien käsityksiin saatetaan yrittää vaikuttaa käsittelemällä erityyppisiä tietoja raportoinnissa vaihtelevin tavoin, esimerkiksi optimistisia arvioita korostamalla (Merkl-Davies & Brennan, 2007). Rutherfordin (2005) mukaan liikevoitosta, kannattavuudesta ja yrityksen varoista puhutaan optimistiseen sävyyn raporttien sanallisissa osuuk-

sisä yhtä paljon riippumatta siitä, onko yrityksen tulos ollut voitollinen vai tappiollinen. Lisäksi Clatworthy ja Jones (2003) havaitsivat, että heikentyviä tuloksia saaneiden yhtiöiden raportoinnissa esiintyy mainintoja huonoista uutisista odotettua vähemmän, mutta positiiviset seikat korostuvat raportoinnissa yrityksen tilasta riippumatta. Toisaalta raportoinnin sävyn on todettu pystyvän ennustamaan yrityksen tulevaa suorituskykyä (Abrahamson & Amir, 1996; Li, 2010), joten vaikka positiivisten uutisten korostamista tai negatiivisten uutisten vähemmälle painotukselle jättämistä raportoinnissa tapahtuisikin, niin sen suuruus ei välttämättä ole käännteentekevää. On mahdollista, että yrityksen informaatioympäristöllä on merkitystä takeettoman positiivisuuden vaikutusten hillitsemisessä, sillä tulosjulkistusten konferenssitilaisuuksien tapauksessa sanallinen informaatio on vahvemmin yhteydessä yrityksen tulevaan suorituskykyyn, mikäli yritystä seuraa usea analyytikko, jotka saattavat vahvistaa yritysjohdon lausunnoissa esiintynyttä positiivisuutta (Matsumoto, Pronk & Roelofsen, 2011).

Tiedon julkaisemisen lisäksi talousraportoinnissa saatetaan antaa syitä raportoiduille tapahtumille tai yhdistää niitä muihin kehityksiin, mikä voi antaa mahdollisuuden sijoittajien harhaanjohtamiseen. Clatworthyn ja Jonesin (2003) mukaan yrityksen heikkenevä suorituskyky yhdistetään raportojen toimesta usein ulkoisiin seikkoihin, kuten toimintaympäristön vaikeuksiin, kun taas onnistuneisiin tuloksiin kerrotaan olevan syynä yrityksen oma kyvykkyys. Lisäksi selityksiä annetaan tyypillisemmin huonoja uutisia julkaistessa (Baginski ym., 2004). Kuitenkin yritykset saattavat välttää yleisten negatiivisten taloudellisten tapahtumien kommentointia, mikäli on mahdollista, että yritystä voitaisiin syyttää kyseisestä tapahtumasta (Schloetzer, Tseng, Yohn & Yoon, 2021), mikä viittäisi Clatworthyn ja Jonesin (2003) huomioon yhdistettynä siihen, että ulkoisia ongelmia saatetaan käyttää huomion poiskohdistamiseen yrityksen omasta tilasta, jos se on ilman mainetappioita toteutettavissa. Toisaalta ulkoisien syiden liittämisen yrityksen tuloksiin on todettu olevan sijoittajien päätöksenteon kannalta hyödyllistä (Baginski ym., 2004), joten vilpillisyyttä ei välttämättä esiinny laajalti tai sijoittajat pystyvät näkemään virheellisen yhdistämisen läpi päätöksiä tehdessään.

Yhteenvetona on todettavissa, että sanallinen sisältö tarjoaa yritykselle mahdollisuuden kattavampaan raportointiin, mutta se on samalla helpommin manipuloitavissa. Kummallekin edellä esitetylle näkökannalle sanallisen informaation luonteesta raportoinnissa, vaikutelmien hallinnoimiselle ja lisätietoa antavalle, on löydetty perusteita. Erilaisia vaikutelmien hallinnoimisen teorioita on todettu käytännössä tapahtuviksi, mutta samaan aikaan on nähty sijoittajien pitävän sanallisessa muodossa annettua informaatiota hyödyllisenä lisänä perinteisen laskentatoimen tuottaman tiedon ohella. Vielä ei ole täysin selvää missä määrin sijoittajat ovat alttiita mahdolliselle yrityksien raportoinnissa esiintyvälle keinottelulle.

3 TEKSTIANALYYSI LASKENTATOIMESSA

3.1 Tekstianalyysi menetelmänä

Tyypillisesti tekstianalyysin tavoitteena laskentatoimen kontekstissa on havaita sanallisessa muodossa olevasta aineistosta mielipiteitä, tunteita tai muita arvo-relevantteja tekijöitä kuvaavaa informaatiota (Marty, Vanstone & Hahn, 2020). Tämä pyritään saavuttamaan erottamalla ja koostamalla kvalitatiivisesta aineistosta tietoa kvantitatiiviseen muotoon, jota voidaan hyödyntää tutkimuksessa tilastollisin menetelmin (Loughran & McDonald, 2016). Laskentatoimen tutkimuksen kontekstissa tekstianalyysia on hyödynnetty vaihtelevin menetelmin ja aineistoin. Aineistot ovat yleensä koostuneet joko yrityksen, median tai muiden ulkopuolisten tahojen, kuten internet-käyttäjien, julkaisemasta materiaalista (Kearney & Liu, 2014).

Tekstianalyysi voidaan nähdä neliosaisena prosessina, johon kuuluu dokumentin jakaminen osiin, aineiston mahdollinen muokkaaminen ja karsiminen, sisällön luokittelu ja lopuksi sen pisteyttäminen. Dokumenttien osittaminen tapahtuu joko sana- tai lausetasolla, kun dokumentit jaetaan pienempiin tarkasteltaviin osiin. Sanakohtainen jakaminen on riittävää silloin, kun sanoja tarkastellaan itsenäisinä, mikä on ollut yleisemmin käytetty lähtökohta tekstianalyysissa laskentatoimen saralla, mutta mikäli halutaan ottaa huomioon sanojen keskinäinen merkitys, voidaan dokumentit jakaa lauseittain. (Marty ym., 2020.) Lausetasoinen jakaminen on ollut harvinaista, sillä niiden luotettava tunnistaminen tekstistä esimerkiksi pisteiden avulla voi olla haastavaa varsinkin talousraporteissa, jotka saattavat sisältää taulukoita, erilaisia numeroita sekä listoja, kun taas sanoittain jakaminen onnistuu kohtuullisella varmuudella välilyönneittäin ja rivinvaihdottain tekstiä pilkkomalla (Loughran & McDonald, 2016). Dokumenttien jakaminen lauseisiin voi kuitenkin tarkentaa analyysia, sillä se mahdollistaa esimerkiksi eri lauseenjäsenien tunnistamisen, joiden roolien avulla lauseiden laatu voidaan arvioida (Marty ym., 2020).

Dokumenttien karsimisen tavoitteena on vähentää niissä esiintyvää tutkimuksen kannalta tarpeetonta tai menetelmän luotettavuutta heikentävää ma-

teriaalia (Marty ym., 2020). Yksi keino tähän on keskittyä vain tiettyihin raporttien osioihin, jotka nähdään oleellisimpina tutkimusongelman kannalta, kuten talousjohdon kommentit tai tilinpäätöksen laatimisperusteet (Loughran & McDonald, 2016). Toisaalta aineistoa voidaan karsia sanoja tarkastelemalla, jolloin saatetaan jättää huomiotta tietyt toistuvat mutta usein merkityksettömät sanat, kuten englannin kielessä artikkelit. Toinen tapa sanojen karsimiseen on sanojen perusmuotoistaminen, jossa sanoista pudotetaan päätteet pois ja jätetään jäljelle vain sanan runko. Tällä tavoitellaan sanojen kategorisoimisen helpottamista, kun jokaista taivutusmuotoa ei tarvitse tunnistaa erikseen. (Marty ym., 2020.) Se voi kuitenkin aiheuttaa väärinkäsityksiä, jos kaksi samantapaista mutta eri merkityksen omaavaa sanaa päätyy tulkittavaksi samoin (Loughran & McDonald, 2011).

Kun aineisto on rajattu ja tarvittaessa muokattu, voidaan sen sisältö kategorisoida, joka on pääasiassa suoritettu erilaisten sanalistojen tai koneoppimisen avulla. Sanalistoina on käytetty sekä yleisiin teksteihin että talousmaailmaan sopivia kirjastoja, joissa sanat on luokiteltu kategorioihin, kuten positiivinen ja negatiivinen. Listojen hyötypuolena on niiden pysyvyys, eli niitä voidaan käyttää uudelleen toisissa tutkimuksissa, jolloin yksittäisen tutkijan subjektiivinen näkemys sanojen laadusta ei vaikuta tuloksiin. Sanalistojen heikkous puolestaan piilee itse menetelmässä, joihin niitä käytetään, sillä dokumentin analysointi sanoja kategorioihin luokittelemalla ei välttämättä kerro koko kuvaa dokumentin laadusta tai sävystä, kun sanojen keskinäiset suhteet jätetään huomioida. (Loughran & McDonald, 2016.) Luotettavampia tuloksia saatetaan saada koneoppimista käyttämällä (Kearney & Liu, 2014). Sen ideana on kouluttaa algoritmi tunnistamaan esimerkiksi lauseen sävy keräämällä ensiksi valmiiksi luokiteltu harjoitusaineisto, jonka jälkeen algoritmi tunnistaa lopun aineiston piirteet tämän perusteella (Marty ym., 2020). Tässä lähestymistavassa haittapuolena on harjoitusaineiston laatimisen vaativuus, sillä se vaatii yleensä manuaalista luokittelua, joka saattaa olla altis virheille (Kearney & Liu, 2014). Lisäksi harjoitusaineistoa voi olla vaikea hyödyntää jatkotutkimuksissa, jolloin tutkimuksen toistettavuus kärsii (Loughran & McDonald, 2016).

Viimeisenä vaiheena tekstianalyyseissa on yleensä dokumentin pisteytys, jossa havainnot saatetaan kvantitatiiviseen muotoon. Koneoppimista hyödynnettäessä aineistoa käsitellään usein lausetasolla, jolloin lopulliseksi mitattavaksi kohteeksi muodostuu esimerkiksi lauseen positiivisuuden todennäköisyys, jonka keskiarvoa mitataan koko dokumentin lauseiden kesken (Kearney & Liu, 2014). Sanalistoihin perustuvassa tutkimuksessa aineisto pisteytetään tavallisesti kategorioihin luokiteltujen sanojen frekvenssien suhteena koko dokumentin sanamäärään (Marty ym., 2020). Pisteytystä saatetaan muokata sanoja painottamalla, jos oletusta sanojen yhtä suuresta merkityksestä ei pidetä tarpeeksi tarkkana (Kearney & Liu, 2014). Esimerkkejä painotuksesta ovat Loughranin ja McDonaldin (2011) käyttämä sanojen yleisyyteen perustuva painotusjärjestelmä, sekä Jegadeeshin ja Wun (2013) muodostama painotus, joka antaa sanoille painon niihin yhdistetyn markkinoiden reaktion voimakkuuden mukaisesti. Lisäksi Boudt ja Thewissen (2019) käyttivät sanojen sijaintiin perustuvaa painotus-

menetelmää, jossa dokumentin alku- ja loppuosassa olleet sanat saivat keski- vaiheilla olevia sanoja korkeamman painotuksen. Pisteytyksen lopputulemana saadaan arvoja, joita voidaan tutkia tilastollisin menetelmin (Kearney & Liu, 2014).

Seuraavaksi käydään läpi aikaisemman tutkimuksen löydöksiä, jotka perustuvat yritysten julkaiseman materiaalin vaikuttavuuteen osaketuottojen kannalta. Tämän jälkeen laajennetaan tarkastelua muihin sisältölähteisiin ja selitettäviin tekijöihin, joihin tekstianalyysia on hyödynnetty laskentatoimen kontekstissa.

3.2 Raportin sisältö ja osakkeen arvo

Yritysten julkaisemien talousraporttien tekstuaalisella sisällöllä on todettu olevan vaikutusta julkaisijan osakkeen arvon muutokseen. Yksi tapa tutkia talousraporttien tekstisisällön vaikutusta on verrata kahta perättäistä raporttia keskenään ja laskea, kuinka paljon ne eroavat toisistaan. Mitä suurempia muutokset edelliseen raporttiin verratessa ovat, sen suurempi vaikuttaisi myös reaktio osakemarkkinoilla olevan (Brown & Tucker, 2011). Suurin osa raporteihin tehdyistä lisäyksistä on negatiivisia, minkä seurauksena muutoksia tehneiden yritysten osakkeen arvon muutos on keskimäärin suhteessa negatiivisempi, kuin sellaisten yritysten, joiden raportit ovat pysyneet samankaltaisina (Cohen ym., 2020). Brownin ja Tuckerin (2011) mukaan yleinen kehitys raporteissa on, että niin muutoksien määrä raporttien välillä kuin myös niiden aiheuttamat osakkeiden hinnanmuutosten suuruudet lyhyellä, muutaman päivän aikavälillä, ovat vähenemään päin. Toisaalta koko seuraavaa vuotta tarkastellessa muutosten seuraukset ovat näkyvämpiä, mikä saattaa johtua sijoittajien kykenemättömyydestä arvostaa talousraporttien tekstissä esiintyviä muutoksia välittömästi (Cohen ym., 2020). Etenkin tekstimuotoinen informaatio voi olla vaikeaa prosessoida ja hyödyntää nopeasti, mikä saattaa hidastaa informaation sisällyttämistä osakkeiden hintoihin (Li, 2006). Lisäksi toisin kuin numeraalinen data, yritysjohdon sanallisista kuvauksista ei yleensä tarjota suoraa vertailua edelliseen raporttiin tai muuhun menneeseen vertailukohtaan, mikä vaikeuttaa kahden julkaistun raportin sanamuotojen vertailemista keskenään entisestään (Cohen ym., 2020).

Raporttien välisten muutosten lisäksi on tutkittu niiden sisällöllisten ominaisuuksien, kuten sävyn, vaikutusta osakemarkkinoilla. Peräkkäisten raporttien välillä tapahtuneet muutokset yritysjohdon näkemyksiä kuvaavan osion positiivisen ja negatiivisen sävyn erotuksessa korreloivat raportin julkaisuhetkellä syntyneiden osaketuottojen kanssa amerikkalaisten yritysten osalta (Feldman, Govindaraj, Livnat & Segal, 2010). Samoin Li, Wu, Xiao ja You (2019) havaitsivat kiinalaisia yrityksiä tarkastellessaan yhteyden positiivisen ja negatiivisen sävyn erotuksen sekä epänormaalien tuottojen välillä lyhyellä, mutta myös pitkällä aikavälillä. Tilinpäätöksissä esiintyvien riskiä kuvaavien sanojen määrän kasvaminen suuresti puolestaan vaikuttaisi olevan yhteydessä heikom-

paan osakkeen arvonkehitykseen seuraavalle 12 kuukaudelle verrattuna vähemmän riskeistä tiedottaneisiin yrityksiin (Li, 2006). Riskiin tai epävarmuuteen verrattavissa olevalla negatiivisella sävyllä näyttäisi olevan vaikutusta osakkeen arvoon lyhyellä aikavälillä tilinpäätöksen julkaisuajankohdan lähetyvillä (Loughran & McDonald, 2011). Negatiivisen sävyn mittaamista pidetään hieman luotettavampana kuin positiivisen, sillä negatiivissävytteisiä sanoja ei juurikaan käytetä ilmaisemaan positiivista asiaa, mutta päinvastainen käytäntö puolestaan on yleisempää (Loughran & McDonald, 2016). Tämän vuoksi yhteys negatiivisen sävyn ja osakkeen arvon vaihtelun välillä on todettu laajemmin. Kuitenkin Jegadeesh ja Wu (2013) löysivät markkinaperusteisella sanojen painotusjärjestelmällään näyttöä sekä positiivisen että negatiivisen sävyn merkittävyydelle. Tämän lisäksi positiivinen sävy näyttäisi olevan olennaista osakkeen arvon kannalta raporttien korjauksien yhteydessä, etenkin jos korjauksissa esiintyy samanaikaisesti suurta varmuutta osoittavaa sävyä (Durnev & Mangen, 2011). Raporttien sävyä on tutkittu myös mittaamalla epänormaalia tai odottamatonta positiivisuutta, joka vaikuttaisi olevan yhteydessä positiivisiin epänormaaleihin tuottoihin (Davis, Piger & Sedor, 2012). Toisaalta tätä positiivista reaktiota saattaa seurata negatiivinen vastareaktio, jos yritysjohton kuvailu on todettu sijoittajien keskuudessa liian optimistiseksi (Huang, Teoh & Zhang, 2014). Edellä mainituissa tutkimuksissa raportin sävyn on todettu olevan merkityksellistä sijoittajien päätöksenteon kannalta, mutta Jiangin, Leen, Martinin ja Zhoun (2019) koostama sävyperusteinen indeksi ei ollut pätevä pitkän aikavälin tuottojen ennustamisessa, vaikka lyhyellä aikavälillä ennustettavuus olikin muiden tutkimusten tulosten tavoin huomattavissa. Yksittäisten yritysten sijaan tämä indeksi kuitenkin oli sopiva koko markkinoiden liikkeiden selittäjänä (Jiang ym., 2019).

Toinen sisällöllisten ominaisuuksien tarkastelukohde on ollut tulevaisuuteen suuntautuvat lauseet tai väitteet talousraporteissa. Sekä numeraalisesti että tekstimuotoisesti annettavat tulevaisuutta koskevat lauseet saavat aikaan reaktion osakemarkkinoilla (Bozanic, Roulstone & Buskirk, 2018). Lisäksi raportin julkaisuhetkellä tapahtuneet osaketuotot vaikuttaisivat vastaavan paremmin tulevien raporttien liikevoiton kehitystä, mikäli tarkasteluhetkellä julkaistussa raportissa on käytetty paljon tulevaisuuteen suuntautuvia lauseita (Muslu, Radhakrishnan, Subramanyam & Lim, 2015). Tulevaisuutta koskevat lausunnot ovat informatiivisia etenkin silloin, kun ne tarjoavat lisätietoa jostakin kvantitatiivisesta tiedosta. Li (2010) havaitsi, että siirtyvien erien merkitys osaketuottojen kannalta vähenee, mikäli yritysjohto on antanut niiden tulkitsemiseen mallittaisia ohjeita tulevaisuutta ajatellen. Lisäksi osakemarkkinoiden reaktio on muihin kategorioihin verrattuna vahvempi, mikäli tulevaa ennakoidaan liikevoiton näkökulmasta (Bozanic ym., 2018).

Allee ja Deangelis (2015) laajensivat raporttien sisällön tarkastelua sana- ja lausetasolla tapahtuvasta frekvenssien tai todennäköisyyksien laskemisesta raportin rakenteen tasolla suoritettavaan tarkasteluun, kun he tutkivat positiivis- ja negatiivissävyisten sanojen jakautumista raportin sisällä. Mikäli informaatio positiivisen tai negatiivisen sävyn välittämänä on jakautunut tasaisesti koko

raportin sisällön kesken, sijoittajien reaktio näyttäisi olevan voimakkaampi. Tästä voi päätellä, että sävyn välittämä informaatio on läpituonkevuudellaan vaikuttavampaa, kun se on ilmaistu koko raportin kattavassa, ehkä jopa narratiivisessa, muodossa (Allee & Deangelis, 2015). Toisaalta, mikäli positiivinen informaatio on sijoitettu raporttien alkuun, sen aiheuttama reaktio osakemarkkinoilla on voimakkaampi (Cheng, Roulstone & van Buns Kirk, 2021). Pelkän sanamuotoisesti esitetyn tiedon määrän lisäksi myös sen sijoittelulla vaikuttaisi siten olevan merkitystä sijoittajien päätöksentekoon.

Vaikka raportit sisältäisivätkin sijoittajille olennaista tekstimuotoista informaatiota, se voi jäädä hyödyntämättä, mikäli sijoittajat eivät pysty tulkitsemaan tai käsittelemään yrityksen lähettämää viestiä. Informaatiojulkaisun monimutkaisuudesta koituvia ongelmia on tutkittu luettavuuden käsitteen avulla. Sanamääriltään pitkät ja siten mahdollisesti vaikeaselkoisemmat tilinpäätösraportit ovat todennäköisemmin yhteydessä sijoittajien alireagointiin kuin lyhyemmät vastaavat (You & Zhang, 2009). Sama tulos on todettu myös neljännesraporttien osalta (Lee, 2012). Vaikeaselkoisia raportteja julkaisevien yritysten osakkeilla käydään vähemmän kauppaa raporttien julkaisuhetkellä, etenkin piensijoittajien toimesta (Miller, 2010). Tämän seurauksena monimutkaisemmat raportit saattavat aiheuttaa epävarmuutta osakemarkkinoilla lyhyellä aikavälillä, mutta niiden informaatio sisältyy kuitenkin lopulta osakkeiden hintoihin (Li & Zhao, 2016). Pitkällä aikavälillä tilinpäätöksen vaikeaselkoisuus ei näyttäisi vaikuttavan osaketuottoihin juurikaan (Li, 2008).

Kokonaisuutena talousraporttien tekstimuotoinen sisältö näyttäisi olevan merkityksellistä sijoittajien päätösten kannalta. Vahvimmin informaatiota vaikuttaisi välittävän negatiivinen sävy, etenkin jos se annetaan tulevaisuuteen suuntautuvassa muodossa. Julkaistujen tietojen sisällymistä osakkeen arvoon voi viivästyä tekstimuotoisen informaation numerodataa hitaampi käsittely ja arvostaminen, minkä seurauksena raportin vaikutukset saatetaan nähdä vasta hiukan sen julkaisemisen jälkeen.

3.3 Muut kontekstit

Toinen tekstianalyysin saralla tärkeä tutkimuskohde osakkeen arvoa selittävien mallien ohella on ollut selvittää, kuinka hyvin julkaistujen raporttien sisällölliset tekijät pystyvät ennustamaan yrityksen tulevaa suorituskykyä laskentatoimen mittarien valossa. Jos raporteissa on kuvailtu tulevia odotuksia positiiviseen sävyyn, yrityksen liikevoitto näyttäisi olevan aiempaa suurempi seuraavien neljän neljänneksen ajan, joskin vaikutus painottuu eniten lähimpään tulevaan neljänneeseen (Li, 2010). Bochkay ja Levine (2019) kuitenkin eivät pystyneet parantamaan liikevoittoa ennustavaa malliaan lisäämällä siihen Lin (2010) tavoin muuttujan, joka perustuu positiivisten ja negatiivisten sanojen määrään raportissa. Sen sijaan painottamalla merkittävimpiä sävykategorioihin kuuluvia sanoja liikevoiton ennustettavuus parani etenkin pienten tai vähällä analytiikkosten seurannalla olevien yritysten osalta (Bochkay & Levine, 2019). Raportin

sävy vaikuttaisi olevan erityisen hyödyllistä liikevoiton ennustettavuuden kannalta silloin, kun yrityksen tulokseen liittyy epävarmuutta esimerkiksi toimialan tai jaksotusten aiheuttamana (Li ym., 2019). Toisaalta positiivinen sävy ennakoi heikompaa tulevaa liikevoiton kehitystä, mikäli raportin positiivisuus on epätavallisen suurta yrityksen tunnuslukuihin ja liiketoimintaympäristön piirteisiin nähden (Huang ym. 2014). Huangin ym. (2014) tulos on mielenkiintoinen verrattuna Davisin ym. (2012) havaintoihin, joiden mukaan sijoittajat reagoivat myös epänormaaliin positiivisuuteen myönteisesti, mistä voisi päätellä, että liikejohdon saattaa olla mahdollista vaikuttaa sijoittajien reaktioon kasvattamalla raportoinnin positiivisuutta, vaikka sille ei välttämättä olisi toiminnallisia perusteita. Epävarmuutta tulosten tulkintaan näiden kahden tutkimuksen välillä kuitenkin aiheuttavat niiden eroavat menetelmät epänormaalin positiivisuuden mittaamiseksi, sillä Davisin ym. (2012) epänormaalius perustui edellisen raportointikerran sävyyn, kun taas Huang ym. (2014) määrittelivät epänormaalin positiivisuuden yrityksen eri suorituskykymittareiden avulla.

Raporttien positiivisen sävyn on todettu pystyvän ennustamaan liikevoiton lisäksi muitakin yrityksen suoritusmittareita, sillä yrityksen kasvava markkina-arvo on yhteydessä positiivisen raportointisävyn kanssa (Xiao Wu, Yao & Luan, 2021). Positiivinen tulevaisuutta kuvaileva sävy voidaan yhdistää myös parantuneeseen likviditeettiin seuraavilla neljänneksillä (Li, 2010). Pessimistinen sävy liikejohdon kommentoissa puolestaan on yhteydessä matalampaan kokonaispääoman tuottoon raportointijaksolla (Davis & Tama-Sweet, 2012). Yrityksen tunnuslukujen ja raportoinnin sävyn yhteyttä on tutkittu myös päinvastaisesta näkökulmasta, jossa sävyä pyritään selittämään erilaisilla yrityksen ominaisuuksilla. Lin (2010) mukaan sävyltään positiivinen raportointi on yleisempää pienempien, tunnusluviltaan hyvätasoisten sekä liikevoiton vaihtelultaan vähäisten yritysten keskuudessa. Varovaisuutta raportointiin pienemmän positiivisuuden muodossa saattaa aiheuttaa suurten yritysten tapauksessa oikeudellinen tai poliittinen tarkkaavaisuus, kun taas toimintaympäristön epävarmuus voi vähentää kasvuyritysten positiivisuutta raportoinnissa (Li, 2010). Lisäksi vaikeassa tai heikkenevässä taloudellisessa asemassa olevat yhtiöt tapavat käyttää raportoinnissaan enemmän negatiivisuutta ja epävarmuutta kuvaavaa sanastoa (Hajek, Olej & Myskova, 2014). Lin (2010) ja Hajekin ym. (2014) tuloksista voidaan päätellä, että vaikka raportoinnin sävy näyttäisikin perustuvan osin yrityksen suorituskykyyn, niin siihen voivat vaikuttaa muutkin seikat, kuten toimintaympäristö ja yrityksen koko.

Yritysten julkaiseman materiaalin ja raportoinnin lisäksi tekstianalyysia on hyödynnetty ulkopuolisten toimijoiden julkaisujen vaikutusten tutkimuksessa. Yhden kokonaisuuden muodostavat analyytikköjen raportointia tarkastelleet tutkimukset. Sijoittajan kannalta merkityksellisen informaation määrän on analyytikoiden julkaisuissa todettu kasvavan tekstimuotoisen tiedon avulla sekä itsenäisesti tarkasteltuna että kvantitatiivista tietoa tukevana (Huang, Zang & Zheng, 2014). Tärkeimmät komponentit sijoittajien päätöksenteon kannalta tekstisisällössä vaikuttaisivat olevan sen sävy ja yksityiskohtaisuus (Twedt & Rees, 2012). Sävy saattaa olla jopa raportin kokoavaa analyytikon

suositusta merkityksellisempi (Huang ym., 2014). Raportin yksityiskohtaisuuden sijoittajat puolestaan vaikuttaisivat tulkitsevan merkinä asiantuntijuudesta (Twedt & Rees, 2012), toisin kuin yritysten julkaisemien raporttien tapauksessa, jossa raportoinnin monimutkaisuutta saatetaan käyttää arvorelevantin informaation peittelemiseen vaikeassa taloudellisessa tilanteessa (Li, 2008).

Toinen tekstianalyysissä käytetty yrityksen ulkopuolinen aineistolähde on ollut uutisointi mediassa. Tetlockin (2007) mukaan uutisoinnin pessimistisyys kasvattaa koko markkinaa tarkastellessa kaupankäynnin määrää hetkellisesti, mutta pessimistisyydestä seurannut negatiivien hinnanmuutos korjautuu lähtötasolle muutaman päivän kuluessa. Yrityskohtaisesti uutisointia tarkastellessa negatiivinen sävy vaikuttaisi laskevan uutisen julkaisun jälkeisiä osaketuottoja sekä ennakoivan heikompaa liikevoittoa (Tetlock, Saar-Tsechansky & Macskassy, 2008). Tetlock ym. (2008) raportoivat osakemarkkinoiden reaktion uutisointiin olevan hiukan viivästynyt, mutta uutisten tarjoama informaatio vaikuttaisi kuitenkin välittyvän osakkeiden hintoihin muutaman päivän aikavälillä, etenkin jos se koskee vaikeammin määriteltävien yrityksen suorituskykymittarien tulkintaa. Sinha (2016) sen sijaan totesi reaktion kestävän huomattavasti pidemmän aikaa ja päätteli siten sijoittajien alireagoivan lyhyellä aikavälillä yrityskohtaisen uutisoinnin sävyyn, mikä tarkoittaisi, että kaikki uutisoinnin informaatio sisältyisi osakkeiden hintaan vasta pidemmällä aikavälillä. Ahmad, Han, Hutson, Kearney ja Liu (2016) puolestaan havaitsivat yrityskohtaisen uutisoinnin seurauksena aiheutuvien reaktioiden olevan kahden tyyppiä, osaksi nopeasti kumoutuvia mutta osaksi myös pidemmän aikavälin suuntaa selittäviä. Pidempää aikaväliä vaativan uutisoinnin informaation osakekursseihin sisällymisen arvellaan johtuvan talousraportoinnin tavoin hitaammasta informaation käsittelystä, kun taas lyhyet ja nopeasti kumoutuvat piikit saattavat johtua sijoittajien ylireagoinnista median korostamiin näkökantoihin (Ahmad ym., 2016).

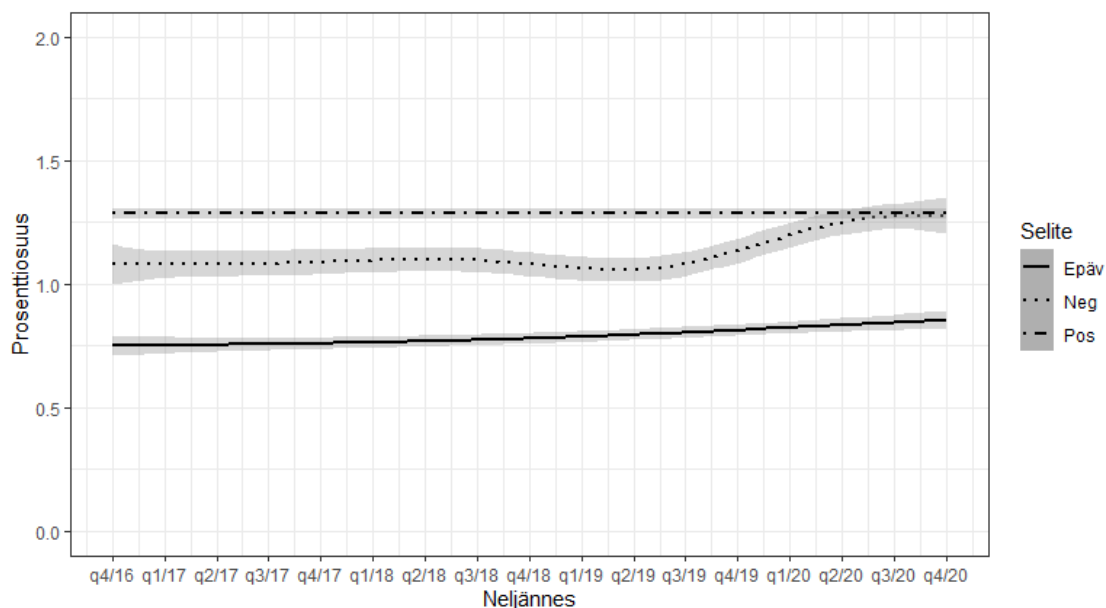
Tekstiaineistoja on hyödynnetty laskentatoimen tutkimuksessa kattavasti sekä erilaisista lähteistä että erilaisiin tarkoituksiin. Osakekurssien vaihtelua on pystytty selittämään myös yrityksen ulkopuolisia tekstilähteitä, esimerkiksi uutisointia, analysoimalla. Lisäksi talousraporttien sisällön on todettu pystyvän ennustamaan yrityksen tulevaa suorituskykyä, jolloin tarkastelua on laajennettu myös pidemmälle aikavälille, toisin kuin tämän tutkimuksen lyhyen aikavälin arvorelevanssianalyysin tapauksessa.

4 AINEISTO JA MENETELMÄT

4.1 Aineisto

Tutkimuksen aineisto koostuu 1118:sta suomalaisten yritysten vuosina 2017–2021 julkaisemasta englanninkielisestä osavuosisiraportista. Aineiston raporttien käsittelemät jaksot sijoittuvat vuoden 2016 viimeisestä neljänneksestä vuoden 2020 viimeiseen neljännekseen. Eri yrityksiä aineistossa on kaiken kaikkiaan 107. Kaikilta yrityksiltä ei aineistossa ole jokaista mahdollista mainitulla aikavälillä julkaistua raporttia joko niiden julkaisemattomuuden tai julkaistun raportin tiedostollisen sopimattomuuden vuoksi. Aineiston raportit on jaoteltu yrityksittäin liitteissä 1 ja 2. Raportit on ladattu pdf-muodossa Nasdaq Nordicin julkaisuarkistosta (Nasdaq Nordic, 2021b) sekä osaksi suoraan yritysten nettisivuilta. Huomionarvoista on, että aineiston tarkasteleman ajankohdan keskivaiheilla käynnistynyt koronavirustilanne on saattanut vaikuttaa yritysten raportointiin. Viitteitä tästä on kuviossa 1, josta on huomattavissa, että varsinkin negatiivisten sanojen keskimääräinen osuus raporteissa vaikuttaisi kääntyneen nousuun vuoden 2019 alusta alkaen. Lisäksi positiivisuudessa on havaittavissa maltillisesti laskeva ja epävarmuudessa hiukan nouseva trendi aineiston loppua kohden. Tämän vuoksi regressiomallien osalta aineistoa on tarkastelu sekä täysimuotoisesti että kahteen osaan vuoden 2019 alusta jakaen. Näin ollen on mahdollista arvioida millä tavalla, jos ollenkaan, poikkeuksellinen koronatilanne on vaikuttanut epänormaalien tuottojen ja raportin tekstuaalisen sävyn väliseen yhteyteen. Lisäksi aineiston jakamisen avulla voidaan selvittää tämän yhteyden luonnetta ns. normaalissa tilanteessa ilman viime aikojen maailmanlaajuisia poikkeusoloja. Kuviossa 1 on nähtävissä myös sävyä kuvaaviksi tunnistettujen sanojen osuuksien kokoluokat, jotka ovat noin yhden prosentin tuntumassa kunkin kategorian osalta, eli suurin osa aineistossa esiintyneistä sanoista on näiden kategorioiden valossa neutraaleja. Sävysanojen osuuksia on eritelty numeraalisessa muodossa taulukossa 3.

Kuvio 1 Sävyvuuttujen keskiarvokäyrät aineistossa



Regressiomalleissa käytettyjen kontrollimuuttujien tiedot on hankittu Refinitiv Eikonin Datastream-palvelusta. Epänormaalien tuottojen määrittämisessä tarvittavien osaketuottojen aineistot on ladattu Nasdaq Nordicin sivuilta jokaisen yrityksen osalta vuoden 2016 alusta vuoden 2021 toukokuuhun (Nasdaq Nordic, 2021a). Samasta lähteestä on hankittu myös koko markkinan kehitystä kuvaamaan OMX Helsinki GI:n (General Index) päivittäinen kehitys. Odotettujen tuottojen laskemiseen vaadittavan kolmen faktorin markkinamallin data puolestaan on ladattu Kenneth Frenchin nettisivuilta Euroopan kolmen faktorin päivittäisessä muodossa (French, 2021).

Nasdaq Nordicin julkaisuarkistosta tehdyt lataukset on suoritettu tietokoneavusteisesti. Arkiston hakukentissä markkinaksi on valittu Nasdaq Helsinki ja julkaisuja on haettu luokista neljännesvuosikatsaus, osavuosikatsaus (Q1 ja Q3), puolivuositiedote sekä tilinpäätöstiedote. Sekä automaattisten että manuaalisten latausten yhteydessä on otettu talteen joko yrityksen tai julkaisuarkiston antama raportin julkaisupäivämäärä, jota käytetään tapahtumatutkimuksen tapahtumapäivänä. Taulukossa 1 on esitelty tutkimuksessa käytetyn aineiston määrä prosessin eri vaiheissa.

Taulukko 1 Aineiston koko eri vaiheissa

Vaihe	Aineiston koko
Kaikki ladatut raportit	1605
Osakkeen hintadataa saatavilla	1555
Kontrollimuuttujien dataa saatavilla	1412
Muutoslaskujen ja puuttuvien havaintojen poistamisen jälkeen	1118

Aikaisemmassa tutkimuksessa aineistona on käytetty amerikkalaisia 10-K ja 10-Q raportteja, jotka pääosin vastaavat sisällöltään esimerkiksi suomalaisia tilinpäätösraportteja, mutta ovat ulkoasultaan pelkistetympiä. Yritysten itse julkaisemat dokumentit sisältävät yleensä grafiikkaa, taulukoita ja kuvia, joten yksi tutkimusprosessin vaiheista on aineistona olevien dokumenttien karsiminen ja niiden sisällön yhtenäistäminen. Esimerkiksi pdf-tiedostoina ladatut dokumentit saa helpohkosti tekstimuotoon, mutta yritysten eroavat käytännöt raporttien laatimisessa, kuten raportissa käytettyjen taulukkojen määrä, saattavat vaatia tarkempaa syyniä dokumenttien tasavertaistamiseksi.

Kukin dokumentti on ladattu pdf-muodossa, josta ne saa melko luotettavasti tekstimuotoon, jossa niitä voidaan käydä läpi sävyä analysoivan ohjelman avulla. Osa ladatuista raporteista ei saanut muutettua tekstimuotoon, joten ne jätettiin pois aineistosta. Ladatut talousraportit noudattavat melko yhtenäistä kaavaa, jossa aluksi kuvaillaan kyseessä ollutta tarkastelujaksoa sanallisesti, minkä jälkeen esitellään taulukkomuodossa esimerkiksi tuloslaskelma ja tase. Niiden jälkeen saattaa olla vielä huomautuksia esimerkiksi tilinpäätösstandardien noudattamisesta. Kiinnostavinta tämän tutkimuksen kannalta on ensimmäinen osio, jossa yritys kuvaa omaa tilannettaan, joten jälkimmäiset osiot on karsittu pois analysoitavasta tekstistä. Se on tehty etsimällä dokumentista jokin taulukko-osioon viittaava termi, kuten tuloslaskelma, joka on omalla rivillään, eli otsikkona, ja poistettu kaikki sen jälkeinen materiaali. Jäljellä olevasta osiosta on poistettu mahdolliset taulukot, alaviitteet ja sivunumerot jakamalla teksti riveihin ja poistamalla niistä sellaiset, jotka eivät vaikuta lauseilta, eli jos niissä on pelkkiä numeroita tai ei lainkaan tyypillisiä lauseisiin kuuluvia sanoja, kuten tiettyjä partikkeleita. Tekstiosuuden karsimista on kuvattu hiukan tarkemmin liitteessä 3.

4.2 Menetelmät

Tutkimuksessa pyritään selittämään epänormaalien tuottojen määrää talousraportin sävyllä. Tämä toteutetaan koostamalla regressiomalleja, joissa selitettävänä muuttujana on kunkin raportin epänormaali tuotto, jota selitetään erilaisilla raportin sävyä kuvaavilla muuttujilla. Näiden lisäksi regressiomalleissa on mukana kontrollimuuttujia, jotka pyrkivät vähentämään ulkopuolisten tekijöiden, kuten yrityksen taloudellisen tilanteen, vaikutusta aineistoa mallintaessa. Regressiomallien ohella tutkimuksessa käytetään kahden otoksen t-testiä positiivisten sanojen lähistöllä esiintyneiden sanojen merkittävyyden tutkimiseen. Seuraavaksi käydään läpi epänormaalien tuottojen määrittämiseen soveltuva tapahtumatutkimuksen menetelmä, jonka jälkeen tarkastellaan selittävinä muuttujina käytettävien sävykategorioiden muodostamista tekstianalyysin avulla. Lopuksi selvennetään positiivisten sanojen lähistöllä olevien sanojen olennaisuuden määrittämiseen käytettyä menetelmää.

4.2.1 Tapahtumatutkimus

Yksi tutkimuksen osa-alueista on tapahtumatutkimuksen suorittaminen, jossa selvitetään talousraportin julkaisuhetkellä syntyneet epänormaalit osaketuotot. Tapahtumatutkimuksen tavoitteena on selvittää, onko jokin tapahtuma aiheuttanut oletettua suurempia eli epänormaaleja muutoksia osakkeen arvoon tapahtumahetkellä. Epänormaalien tuottojen määrittämiseksi tarvitaan markkinamalli, joka kuvaa odotettavissa olevia tuottoja kullakin aikavälillä. Odotetut tuotot perustuvat tapahtumaa edeltävien pörssipäivien kehitykseen, jota kutsutaan estimointi-ikkunaksi, jonka pituus on tässä tutkimuksessa 252 pörssipäivää. (Benninga, 2014.) Tässä tutkimuksessa odotetut tuotot määritellään Capital Asset Pricing (CAP) -malliin tai Faman ja Frenchin kolmen faktorin malliin perustuen. CAP-malli selittää yrityksen osakkeen arvon vaihtelua koko markkinan kehitykseen perustuen, kun taas kolmen faktorin malli ottaa huomioon koko markkinan muutosten lisäksi yrityksen koon sekä eron kasvu- ja arvoyritysten välillä (ks. esim. Fama & French, 1993). Epänormaalit tuotot saadaan vähentämällä julkaisupäivän aikaikkunan tuotoista markkinamallin perusteella muodostetut odotetut tuotot (Benninga, 2014). Tässä tutkimuksessa käytetään 10 päivän tarkasteluväliä sekä ennen raportin julkaisua että sen jälkeen, jota voidaan myös jakaa pienempiin osiin analyysin tueksi. Lyhyen aikavälin tuottoja on tarkasteltu tekstianalyysin yhteydessä yleensä muutaman päivän aikaikkunalla (ks. esim. Feldman ym., 2010), mutta tekstimuotoisen informaation sisäistäminen voi viedä sijoittajilta enemmän aikaa (ks. esim. Cohen ym., 2020), joten 10 päivän aikaväli saattaa tavoittaa paremmin informaation hitaamman sisällyttymisen osakkeiden hintoihin. Epänormaalit tuotot määritetään aikaikkunan väliltä kumulatiivisena summana. Seuraavaksi esitetään tarkemmin epänormaalien tuottojen määrittämisen prosessi.

Kuten mainittua, odotetut tuotot määritetään tapahtumakohtaisesti joko CAP-mallin tai Faman ja Frenchin kolmen faktorin mallin avulla. CAP-mallissa arvopaperin odotettu tuotto perustuu sen kovarianssiin koko markkinan kehityksen kanssa, ja sitä pidetään yksinkertaisuudessaan käyttökelpoisena odotetun tuoton määrittäjänä (MacKinlay, 1997). Olennainen tekijä CAP-mallissa on yrityskohtainen kerroin β , joka kuvaa osakkeen arvon heilahteluallttiutta koko markkinaan nähden. Mikäli β saa arvon 1, osakkeen hinta vaihtelee markkinan kanssa samaan tahtiin, kun taas pienempi (suurempi) arvo tarkoittaa markkinaa pienempää (suurempaa) vaihtelua. (Sharpe, 1964.) Odotetut tuotot CAP-mallin avulla määritetään seuraavasti.

$$E(R_i) = R_f + \beta_i(R_m - R_f) + \alpha \quad (\text{kaava 1})$$

Kaavassa 1 $E(R_i)$ on yrityksen i odotettu tuotto, R_f on riskivapaa tuotto, β_i on riskikerroin, R_m on koko markkinan tuotto sekä α on virhetermi.

Toinen vaihtoehto odotettujen tuottojen määrittämiseen on Faman ja Frenchin faktorimalli, jossa on kolme osatekijää. Osatekijät perustuvat koko markkinoiden tuottoihin riskittömällä tuotolla vähennettynä, pienten ja suurten yritys-

ten tuottojen eroon (SMB) sekä arvo- ja kasvuosakkeiden välisiin tuottoeroihin (HML). Arvo-osakkeilla tarkoitetaan sellaisten yritysten osakkeita, joiden kirjainpidollinen arvo on suurempi kuin markkina-arvo eli joiden P/B-luku (*price-to-book ratio*) on pieni. Tarkemmin sanottuna HML on arvo-osakkeista muodostetun portfolion tuotot vähennettynä kasvuosakeportfolion tuotoilla, kun taas SMB on pienten yritysten portfolion tuotot suuryritysten portfolion tuotoilla vähennettynä.

$$E(R_i) = R_f + \beta_{mi}(R_m - R_f) + \beta_{ki}SMB + \beta_{ai}HML + \alpha \quad (\text{kaava 2})$$

Kaavassa 2 β_{mi} on markkinan kehitystä kuvaavan muuttujan kerroin, β_{ki} on koon vaikutuksen mittaava kerroin sekä β_{ai} on arvoyritysten kehitystä kuvaavan muuttujan kerroin. (Fama & French, 1993.)

Malli on siten laajennus CAP-mallista, sillä se ottaa huomioon sekä pienten yritysten että arvo-osakkeiden tyypillisen koko markkinaa paremman kehityksen, mikä saattaa parantaa osakkeen arvon kehityksen ennustettavuutta CAP-malliin verrattuna. Kullakin osatekijällä on oma kertoimensa, joka määrittyy sekä osakkeen että osatekijöiden estimointi-ikkunan ajalta olevaan historialliseen dataan perustuen, kun osakkeen arvoa regressoidaan osatekijöiden yhteisvaikutuksella. Kyseessä on siis lineaarinen regressiomalli, jossa yrityksen osakkeen tuottoja selitetään kolmen osatekijän kehityksellä. (Fama & French, 1993.)

Griffinin (2002) mukaan kolmen faktorin mallin suorituskyky paranee, jos siinä käytetään maakohtaista dataa faktoreiden määrittämiseen koko maailmaa koskevan aineiston sijasta. Sen vuoksi tässä tutkimuksessa saatetaan saada tarkempia tuloksia käyttämällä markkinadatana Helsingin pörssin yleisindeksin (OMX Helsinki GI) päivittäistä kehitystä. Kahden muun faktorin osalta aineistoa on saatavilla päivittäisenä koko Eurooppaa koskien. Tässä tapauksessa siis SMB mallintaa sitä, millä tavalla pienet ja kasvavat yritykset suoriutuvat Euroopan alueella, kun taas HML kertoo vakaampien yritysten arvostuksesta Euroopassa. Aineistossa olevien yritysten osalta tarkastellaan niiden vertautuvuutta Euroopan tasolla faktoriportfolioiden tuottoihin, mikä näkyy faktoreiden kertoimien arvossa. Positiivinen kerroin viittaa siihen, että yrityksen osakkeen hinnanvaihtelut vastaavat faktorin kuvaaman ryhmän vaihtelua, eli esimerkiksi positiivinen SMB-kerroin kertoo osakkeen hinnanmuutosten noudattavan jossain määrin pienten yritysten tyypillistä vaihtelua Euroopan alueella suuryrityksiin verrattuna. Negatiivisen kertoimen saaneiden yritysten tuotot vaihtelevat päinvastoin kuin faktorin kuvaama portfolio, ja nollassa lähellä oleva kerroin kertoo yhteyden puutteesta yrityksen osakkeen sekä faktoriportfolion hinnanvaihteluiden välillä.

Kuitenkaan kaikkien yritysten osalta Euroopan tasolla laadittu kahden faktorin data ei välttämättä paranna mallin ennustavuutta, vaan päinvastoin voi heikentää sitä, mikäli faktoridata korreloi vähäisesti jonkin osakkeen hinnanmuutosten kanssa. Sen vuoksi tässä tutkimuksessa lasketaan aluksi jokaisen raportin tapauksessa odotettujen tuottojen malli estimointi-ikkunan ajalta sekä

CAP- että kolmen faktorin mallin avulla. Näistä kahdesta valitaan tapahtumaikkunan aikana sovellettavaksi korkeamman selityksasteen (R^2) saanut. Malleja vertailtaessa on syytä käyttää korjattua selityksastetta, sillä muuttujien lisääminen kasvattaa mallin selityksastetta, mutta korjattu selityksaste voi olla muuttujien lisäämisen jälkeen pienempi, jos uudet muuttujat, tässä tapauksessa SMB ja HML, eivät paranna mallin ennustettavuutta tarpeeksi (Kaakinen & Eltonen, 2021).

Kun estimointi-ikkunan tiedoilla muodostetuista regressiomalleista on valittu selityksasteeltaan parempi, saadaan mallista muuttujien kertoimet. Niiden avulla kootaan tapahtumapäivien odotetut tuotot, kun toteutuneet havainnot kultakin päivältä kerrotaan omalla kertoimellaan. Epänormaalit tuotot puolestaan saadaan vähentämällä tapahtumapäivän toteutuneista tuotoista odotetut tuotot. Päivittäinen riskivapaa tuotto sekä mallin virhetermi ovat oletettu nol-laksi, joten ne on poistettu kaavasta.

$$AR_{it} = R_{it} - \hat{\beta}_{mi}R_{mt} + \hat{\beta}_{ki}SMB_t + \hat{\beta}_{ai}HML_t \quad (\text{kaava 3})$$

Kaavassa 3 AR_{it} on yrityksen i ajanhetken t epänormaali tuotto, R_{it} on toteutu-nut tuotto, $\hat{\beta}_{xi}$ on kunkin muuttujan estimoitu kerroin, R_{mt} on markkinaindek-sin tuotto sekä SMB_t ja HML_t faktoriportfolioiden tuotot. (MacKinlay, 1997.)

Selittävänä muuttujana tämän tutkimuksen regressiomalleissa käytetään aikaikkunan epänormaaleja tuottoja, jotka saadaan laskemalla kumulatiivinen summa päivittäisistä epänormaaleista tuotoista.

$$CAR_i(t_1, t_2) = \sum_{t=t_1}^{t_2} AR_{it} \quad (\text{kaava 4})$$

Kaavassa 4 $CAR_i(t_1, t_2)$ on aikaikkunan $[t_1, t_2]$ kumulatiivinen epänormaali tuot-to yrityksellä i . (MacKinlay, 1997.)

Epänormaalien tuottojen määitys on tässä tutkimuksessa tehty tietoko-neavusteisesti, siitä tarkemmin liitteessä 3.

4.2.2 Tekstianalyysi

Tapahtumatutkimuksen tuloksena saatuja epänormaaleja tuottoja selitetään tekstianalyysin avulla muodostetuilla sävykategorioilla. Analyysin apuna käytetään sanalista, jonka avulla kunkin dokumentin sisältö voidaan kategorisoi-da. Aikaisemmissa tutkimuksissa on käytetty sekä yleisiä että talousajatteluun sopivia kirjastoja, joiden avulla tekstin sisältöä voidaan luokitella. Tähän tutki-mukseen sanalistaksi on valittu Loughranin ja McDonaldin (2011) esittelemä lista, jossa sanoja on luokiteltu mm. positiivisuuden, negatiivisuuden ja epä-varmuuden mukaisesti, mutta ennen kaikkea luokittelussa on otettu huomioon sanojen merkitys taloudellista tilaa kuvaavina. Esimerkiksi englannin kielen sana liability voitaisiin tavallisesti mieltää negatiiviseksi, vaikka liiketoimintaa kuvaillessa sen merkitys on melko neutraali. Sanalistassa on yhteensä yli 86 000 luokiteltua sanaa, joista 354 on positiivisia, 2355 negatiivisia ja 297 epävarmuut-

ta kuvaavia. Esimerkkejä kustakin sanaluokasta on liitteessä 3. Kirjasto on vapaasti saatavilla, ja sen lisäksi samasta yhteydestä voi ladata Python-ohjelman, joka ryhmittelee annetun dokumentin sanat kirjaston mukaisesti (Loughran & McDonald, 2020). Ohjelmaa on hiukan muokattu tätä tutkimusta varten esimerkiksi painotettujen kategorioiden laskemista varten. Listan sanat ovat englanniksi, joten valittujen yritysten raportit on ladattu englanninkielisinä, jolloin tarkastelun kohteeksi valikoituvat luonnollisemmin suuryritykset, joiden liiketoiminta on kansainvälistä.

Dokumentit tulee muuttaa tekstitiedostoiksi, jotta niiden sisältöä voidaan analysoida Loughranin ja McDonaldin ohjelman avulla. Ohjelma tulostaa lopputuloksina muutaman eri kategorian prosenttiosuuksia, eli kuinka monta prosenttia kunkin kategorian sanat muodostavat koko dokumentin sanamäärästä. Näitä prosenttiosuuksia käytetään muuttujina regressioanalyysissä. Seuraavaksi käydään läpi malleissa käytettyjen muuttujien määritelmät sekä itse regressiomallit.

4.2.3 Muuttujat ja regressiomallit

Tekstianalyysin perusteella syntyy muuttujia, joiden yhteisvaikutusta voidaan tutkia monimuuttujaregressiomallin avulla. Muuttujien avulla voidaan muodostaa malleja ja tutkia, kuinka hyvin mikäkin yhdistelmä selittää epänormaalien tuottojen vaihtelua. Ensimmäinen malli koostuu kunkin kategorian prosenttiosuuksia kuvaavista muuttujista. Kategorioiden selittävyttä voi olla mahdollista parantaa niiden sanoja painottamalla, mitä hyödynnetään mallissa 2. Loughranin ja McDonaldin (2011) mukaan pelkkiä prosenttiosuuksia käsitellessä oletetaan, että jokainen sana on yhtä merkityksellinen informaation välittämisen kannalta. Kuitenkaan näin ei aina ole, ja yksi lähestymistapa tämän ongelman ratkaisemiseksi on olettaa, että harvinaisemmin esiintyvät sanat ovat poikkeuksellisuudellaan tarkemmin dokumentin sävyä kuvaavia. Eräs keino tämänkaltaiseen painottamiseen on käänteisen esiintymistiheyden mukainen painotus, jossa kunkin termin paino riippuu siitä, kuinka usein se esiintyy koko aineistossa. Sen seurauksena harvemmin koko aineistossa esiintyvät sanat saavat enemmän painoarvoa. (Loughran & McDonald, 2011.) Painottamisen kaava on seuraava:

$$w_{i,j} = \begin{cases} \frac{(1+\log(tf_{i,j}))}{(1+\log(a_j))} \log \frac{N}{df_i} & , \text{jos } tf_{i,j} \geq 1 \\ 0 & , \text{muuten} \end{cases} \quad (\text{kaava 5})$$

, jossa i = sana, j = dokumentti, w = sanan paino, tf = sanan i frekvenssi dokumentissa j , a = dokumentin j sanamäärä, N = aineiston koko, df = kuinka monessa muussa dokumentissa sana i esiintyy.

Viimeinen regressiomalli sisältää eri sävykategorioiden muutoksen saman yrityksen havaintojen välillä. Feldman ym. (2010) esittävät perusteluita muutosten mittaamiseen suorien prosenttiosuuksien sijaan. Ensinnäkin saman yrityksen raportit noudattavat usein samaa pohjaa, jota muokataan raportointikerto-

jen välillä. Tämän seurauksena sävymuuttujissa saattaa esiintyä autokorrelaatiota, mitä voidaan vähentää muutoksia seuraamalla. Toiseksi, sanalistoja käytettäessä sävyn määrittämiseen jotkin sanat voivat tulla toistuvasti väärin luokitelluksi, jos ne kuuluvat johonkin yrityksen toimintaan kiinteästi kuuluvaan osa-alueeseen, mutta eivät välttämättä kuvaa raportin sävyä, esimerkiksi negatiiviseksi luokiteltu sana "waste" jätealan yrityksillä. Muutoksia mittaamalla tästä ongelmasta saatetaan päästä eroon, jos virheellisiä luokitteluja tapahtuu kutakuinkin saman verran yhden yrityksen raporttien välillä, jolloin muutosta kuvaavaan havaintoon jäisivät aidot erot raporttien sävyssä (Feldman ym., 2010).

Muiden selittävien tekijöiden huomioonottamiseksi regressiomalliin tulee lisätä kontrollimuuttujia, jotka kuvaavat muita todennäköisiä syitä epänormaaleille tuotoille. Lisäämällä kontrollimuuttujia regressiomalleihin voidaan yksittäisen sävymuuttujan yhteyttä epänormaaleihin tuottoihin tarkastella sellaisessa tilanteessa, jossa muut mahdolliset osakkeen arvoon vaikuttavat tekijät ovat vakioituja. Kontrollimuuttujiksi tähän tutkimukseen on valittu osakekohtainen tulos jaettuna osakkeen hinnalla, liikevoittoprosentin muutos, velan ja oman pääoman suhteen muutos sekä current ration muutos. Malleissa käytetyt muuttujat on koottu selityksineen taulukkoon 2 ja muuttujien ominaisuudet taulukkoon 3.

Taulukko 2 Regressiomalleissa käytetyt muuttujat

<i>Sävymuuttujat</i>	<i>Selitys</i>	
<i>Pos</i>	Positiivisten sanojen osuus koko dokumentin sanamäärästä prosentteina	
<i>Neg</i>	Sama kuin yllä negatiivisille sanoille	
<i>Epäv</i>	Sama kuin yllä epävarmuutta kuvaaville sanoille	
<i>Pos_p, Neg_p, Epäv_p</i>	Sävymuuttujien painotettu osuus kaavan 5 mukaisesti	
<i>Pos_m, Neg_m, Epäv_m</i>	Prosenttiosuuden muutos yksiköinä saman yrityksen perättäisten havaintojen osalta	
<i>Kontrollimuuttujat</i>	<i>Selitys</i>	<i>Eikon-koodi</i>
<i>EPS/hinta</i>	Osakekohtainen tulos jaettuna raportin julkaisupäivän osakkeen hinnalla	TR.BasicNormalizedEps
<i>Liikevoitto_m</i>	Liikevoittoprosentin muutos yksiköinä	TR.NetProfitMargin
<i>VOPO_m</i>	Velan ja oman pääoman suhdeprosentin muutos yksiköinä	TR.TtlDebtToTtlEquityPct
<i>CR_m</i>	Current ration muutos yksiköinä	TR.CurrentRatio

Taulukko 3 Aineiston ominaisuuksia. Otokoko 1118.

<i>Muuttuja</i>	<i>10% alle</i>	<i>Keskiarvo</i>	<i>90% yli</i>	<i>Keskihajonta</i>
<i>Sanamäärä</i>	2006	4467.71	7298	2152.37
<i>Positiivisuus</i>	0.90	1.44	2.08	0.49
<i>Negatiivisuus</i>	0.58	1.23	1.96	0.53
<i>Epävarmuus</i>	0.43	0.79	1.21	0.33
<i>Pos_p</i>	2.07	5.05	8.62	2.62
<i>Neg_p</i>	2.54	8.94	16.49	5.67
<i>Epäv_p</i>	1.05	2.90	5.21	1.64
<i>Pos_m</i>	- 0.43	- 0.004	0.43	0.38
<i>Neg_m</i>	- 0.39	0.02	0.42	0.35
<i>Epäv_m</i>	- 0.24	0.004	0.25	0.23
<i>EPS/hinta</i>	- 0.03	0.14	0.032	1.87
<i>Liikevoitto_m</i>	- 10.02	- 1.32	8.17	37.74
<i>VOPO_m</i>	- 11.73	1.93	19.09	59.56
<i>CR_m</i>	- 0.25	- 0.008	0.24	0.37
<i>CAR [-10, 9]</i>	- 10.90	0.07	12.31	10.48
<i>CAR [0, 1]</i>	- 6.88	0.08	8.18	6.50
<i>CAR [0, 4]</i>	- 8.68	- 0.17	8.49	7.71
<i>CAR [0, 9]</i>	- 10.59	- 0.82	9.41	8.98
<i>CAR [5, 9]</i>	- 4.74	- 0.64	3.48	4.73

Muuttujien välisiä korrelaatioita on kuvattu taulukossa 4. Suuren otoskoon vuoksi tilastollisen merkitsevyyden raja on melko alhainen, 95% noin 0.059, mutta kaikki muuttujat eivät kuitenkaan korreloi keskenään tilastollisesti merkitsevällä tasolla. Sävyuuttujista positiivisuus korreloi negatiivisesti epävarmuuden kanssa, kun taas negatiivisuuden kasvaessa myös epävarmuus vaikuttaisi kasvavan, ja päinvastoin. Tämä havainto saattaa selittyä sillä, että negatiivisia ilmauksia pehmennetään epävarmoilla sanoilla, mutta positiiviset seikat pyritään tuomaan esiin varmempina. Tämä ilmiö heijastuu myös toisiin sävyuuttujatyyppeihin, esimerkiksi negatiivisuuden ja epävarmuuden muutokset korreloivat keskenään positiivisesti. Painotetun sävyn muuttujat korreloivat positiivisesti keskenään melko vahvasti, mihin syynä voi olla, että tietyt yritykset käyttävät poikkeuksellista sanastoa raportoinnissaan, mikä sitten näkyy kaikissa sävyuuttujissa, sillä painotettu muuttuja saa korkeampia arvoja, mikäli tunnistettu sävysana on aineistossa harvinainen.

Epänormaalien tuottojen osalta negatiivisuus korreloi negatiivisesti kaikilla muuttujatyypeillä jollakin aikajaksolla, mutta positiivisuudella ja epävarmuudella havaintoja tilastollisesti merkitsevästä korrelaatiosta epänormaalien tuottojen kanssa ei juuri ole. Kontrollimuuttujista liikevoittoprosentin muutoksella sekä velan ja oman pääoman suhteen muutoksella on havaittavissa korrelaatiota epänormaalien tuottojen kanssa usealla ajanjaksolla. Mielenkiintoinen havainto on myös positiivisuuden sekä velan ja oman pääoman suhteen välinen negatiivinen korrelaatio, vaikkakin sen tarkasteleminen menee tämän tutkimuksen näkökulman ulkopuolelle.

Taulukko 4 Muuttujien välisiä korrelaatioita. Otokoko 1118. 95% merkitsevyytasolla olevat merkitty tähdellä (*).

	<i>pos</i>	<i>neg</i>	<i>epäv</i>	<i>pos_m</i>	<i>neg_m</i>	<i>epäv_m</i>	<i>pos_p</i>	<i>neg_p</i>	<i>epäv_p</i>
<i>pos</i>	1.000	0.021	-0.068*	0.366*	-0.056	-0.039	0.340*	-0.105*	-0.148*
<i>neg</i>	0.021	1.000	0.121*	-0.047	0.339*	0.079*	0.073*	0.533*	0.202*
<i>epäv</i>	-0.068*	0.121*	1.000	-0.047	0.104*	0.343*	0.112*	0.219*	0.446*
<i>pos_m</i>	0.366*	-0.047	-0.047	1.000	-0.114*	-0.107*	0.095*	-0.060*	-0.058
<i>neg_m</i>	-0.056	0.339*	0.104*	-0.114*	1.000	0.261*	-0.038	0.123*	0.050
<i>epäv_m</i>	-0.039	0.079*	0.343*	-0.107*	0.261*	1.000	0.036	0.095*	0.158*
<i>pos_p</i>	0.340*	0.073*	0.112*	0.095*	-0.038	0.036	1.000	0.536*	0.506*
<i>neg_p</i>	-0.105*	0.533*	0.219*	-0.060*	0.123*	0.095*	0.536*	1.000	0.669*
<i>epäv_p</i>	-0.148*	0.202*	0.446*	-0.058	0.050	0.158*	0.506*	0.669*	1.000
<i>eps/hinta</i>	-0.012	-0.025	0.017	-0.010	-0.026	-0.019	-0.041	-0.059*	-0.030
<i>CR_m</i>	0.037	0.005	0.018	0.027	-0.020	0.019	0.026	0.012	0.014
<i>liikev%_m</i>	0.052	-0.041	-0.011	0.046	-0.041	-0.040	0.050	-0.017	0.015
<i>vopo_m</i>	-0.061*	0.030	0.014	-0.027	0.049	0.005	-0.032	0.016	0.015
<i>CAR[0,1]</i>	-0.024	-0.104*	0.042	-0.016	-0.117*	-0.024	-0.023	-0.065*	0.003
<i>CAR[-10,9]</i>	0.031	-0.139*	0.018	0.053	-0.104*	-0.039	0.016	-0.071*	-0.022
<i>CAR[0,4]</i>	-0.014	-0.068*	0.023	-0.031	-0.084*	-0.011	-0.038	-0.057	-0.025
<i>CAR[0,9]</i>	-0.002	-0.068*	0.020	0.017	-0.084*	-0.029	-0.016	-0.036	-0.012
<i>CAR[5,9]</i>	0.019	-0.017	0.001	0.083*	-0.024	-0.038	0.031	0.023	0.017
	<i>eps/hinta</i>	<i>CR_m</i>	<i>Liikev%_m</i>	<i>vopo_m</i>	<i>CAR[0,1]</i>	<i>CAR[-10,9]</i>	<i>CAR[0,4]</i>	<i>CAR[0,9]</i>	<i>CAR[5,9]</i>
<i>pos</i>	-0.012	0.037	0.052	-0.061*	-0.024	0.031	-0.014	-0.002	0.019
<i>neg</i>	-0.025	0.005	-0.041	0.03	-0.104*	-0.139*	-0.068*	-0.068*	-0.017
<i>epäv</i>	0.017	0.018	-0.011	0.014	0.042	0.018	0.023	0.02	0.001
<i>pos_m</i>	-0.01	0.027	0.046	-0.027	-0.016	0.053	-0.031	0.017	0.083*
<i>neg_m</i>	-0.026	-0.02	-0.041	0.049	-0.117*	-0.104*	-0.084*	-0.084*	-0.024
<i>epäv_m</i>	-0.019	0.019	-0.04	0.005	-0.024	-0.039	-0.011	-0.029	-0.038
<i>pos_p</i>	-0.041	0.026	0.05	-0.032	-0.023	0.016	-0.038	-0.016	0.031
<i>neg_p</i>	-0.059*	0.012	-0.017	0.016	-0.065*	-0.071*	-0.057	-0.036	0.023
<i>epäv_p</i>	-0.03	0.014	0.015	0.015	0.003	-0.022	-0.025	-0.012	0.017
<i>eps/hinta</i>	1	-0.003	0.021	-0.019	0.04	0.034	0.027	0.015	-0.016
<i>CR_m</i>	-0.003	1	0.075*	-0.095*	0	0.007	0.006	0.015	0.019
<i>liikev%_m</i>	0.021	0.075*	1	-0.047	0.093*	0.12	0.096*	0.106*	0.044
<i>vopo_m</i>	-0.019	-0.095*	-0.047	1	-0.048	-0.099*	-0.082*	-0.096*	-0.049
<i>CAR[0,1]</i>	0.04	0	0.093*	-0.048	1	0.642*	0.864*	0.749*	0.015
<i>CAR[-10,9]</i>	0.034	0.007	0.12*	-0.099*	0.642*	1	0.74*	0.862*	0.431*
<i>CAR[0,4]</i>	0.027	0.006	0.096*	-0.082*	0.864*	0.74*	1	0.85*	-0.016
<i>CAR[0,9]</i>	0.015	0.015	0.106*	-0.096*	0.749*	0.862*	0.85*	1	0.513*
<i>CAR[5,9]</i>	-0.016	0.019	0.044	-0.049	0.015	0.431*	-0.016	0.513*	1

Edellä esitetyistä muuttujista on muodostettu kolme regressiomallia ensimmäisen tutkimuskysymyksen pohjalta. Saman sävymuuttujan eri muotoja, kuten positiivisuus ja positiivisuuden muutos, ei haluttu samaan malliin, jotta voitaisiin tarkastella jokaista muuttujaryhmää omana tapauksenaan. Yksi ryhmä voidaan nähdä yhtenä tapana kuvata raportin sävyä, joten jakamalla muuttajat kolmeen malliin eri ryhmien osuvuutta voidaan vertailla esimerkiksi selityksasteen kautta. Lisäksi saman sävymuuttujan eri muotojen välillä on havaittavissa korrelaatiota (ks. taulukko 4), joten kaikkien muuttujien lisääminen samaan malliin saattaisi saada aikaan multikollinearisuutta ja heikentää siten mallin tulkittavuutta. Toisaalta painotettujen muuttujien välillä on myös huomattavissa kohtalaista korrelaatiota, mikä voi heikentää niitä käyttävän mallin luotettavuutta toisiin malleihin verrattuna. Jokaisessa mallissa selitettävänä muuttujana on raportin julkaisuhetken lähistöllä syntyneet epänormaalit tuotot eri ajanjaksoilla.

Malli 1: Sävy ja kumulatiivinen epänormaali tuotto kontrollimuuttujineen

$$CAR = \beta_0 + \beta_1 * Pos + \beta_2 * Neg + \beta_3 * Epäv + \beta_4 * EPS\hint a + \beta_5 * Liikevoitto_m + \beta_6 * VOPO_m + \beta_7 * CR_m + \varepsilon$$

Malli 2: Painotettu sävy ja kumulatiivinen epänormaali tuotto kontrollimuuttujineen

$$CAR = \beta_0 + \beta_1 * Pos_p + \beta_2 * Neg_p + \beta_3 * Epäv_p + \beta_4 * EPS\hint a + \beta_5 * Liikevoitto_m + \beta_6 * VOPO_m + \beta_7 * CR_m + \varepsilon$$

Malli 3: Sävyyn muutos ja kumulatiivinen epänormaali tuotto kontrollimuuttujineen

$$CAR = \beta_0 + \beta_1 * Pos_m + \beta_2 * Neg_m + \beta_3 * Epäv_m + \beta_4 * EPS\hint a + \beta_5 * Liikevoitto_m + \beta_6 * VOPO_m + \beta_7 * CR_m + \varepsilon$$

Näissä malleissa regressiokertoimien ja samalla koko mallin tulkinta on samankaltaista tilastollisen merkitsevyyden osalta. Jos ne eroavat merkitsevästi nolasta, selitettävän ja selittävän muuttujan välillä on yhteyttä. Yksittäisten muuttujien ohella koko mallin tilastollinen merkitsevyys voidaan määrittää. Lisäksi olennaista on mallien selityksaste, joka kertoo kuinka paljon selitettävän muuttujan vaihtelusta voidaan selittää muiden muuttujien vaihtelulla. Mallien vertaamiseen käytetään korjattua selityksastetta, joka ottaa huomioon selitettävien muuttujien lukumäärän. (Kaakinen & Ellonen, 2021.) Regressioanalyysi toteutetaan tässä tutkimuksessa R-ohjelmistolla.

4.2.4 Positiivisten sanojen lähisanat

Tutkimuksessa tarkastellaan myös positiiviseksi luokiteltujen sanojen lähistöllä esiintyviä sanoja, jotta voitaisiin päätellä, mitkä aihepiirit positiivisten sanojen yhteydessä muuttavat sijoittajien reaktiota. Tällä tavoin esimerkiksi sanaparien "kasvanut tuotto" ja "kasvaneet kustannukset" ero voitaisiin huomata, sillä

pelkän sanan "kasvaminen" huomioinen antaa molemmille termeille positiivisen merkinnän. Jokaisesta raportista tallennetaan kaikki positiivisesta sanasta maksimissaan viiden sanan päässä olevat sanat tai vähemmän, mikäli lause loppuu tai alkaa ennen viiden täyttymistä. Kunkin tallennetun sanan kohdalla raporttiaineisto jaetaan kahtia sen mukaan, onko sana esiintynyt positiivisen termin yhteydessä vai ei. Näiden kahden ryhmän osalta lasketaan keskimääräinen kumulatiivinen epänormaali tuotto tapahtumapäivien 0-4 osalta. Aikaväli on lyhyempi kuin koko raportin sävyä tarkasteltaessa, sillä nyt tarkastelun kohteena ovat yksittäiset sanat koko raportin informaation sijasta, joten saanaan liittyvä informaation voi olettaa olevan nopeampi käsitellä ja tämän pitäisi siten näkyä myös nopeampana reaktiona. Keskiarvojen yhtäsuuruutta testataan kahden otoksen t-testillä, ja mikäli testisuure saa tarpeeksi paljon nollasta eroavan arvon, kyseisen sanan esiintyessä positiivisen termin yhteydessä voidaan todeta muutos keskimääräisissä epänormaaleissa tuotoissa.

Keskiarvotestin jälkeen saadaan tuloksena lista sanoja, joilla näyttäisi olevan yhdessä positiivisen sävyn kanssa vaikutusta sijoittajien reaktioon lyhyellä aikavälillä. Tätä sanalista analysoidaan laadullisesti aineistoa teemoittelemalla. Teemoittelu on yksi laadullisen sisällönanalyysin alalajeista, jossa pyritään yhdistelemään aineistossa olevia elementtejä yhtenäisiksi teemoiksi (Eskola & Suoranta, 1998). Vastaavan kaltaista kategoriointia on esiintynyt myös aikaisemmassa tekstianalyysin tutkimuksessa. Li (2010) käytti 12 kategoriaa tulevaisuutta kuvaavien lauseiden luokitteluun, kun taas Bozanic ym. (2018) hyödynsivät muutaman sanan listaa liikevoittoa koskevien lauseiden määrittämiseen. Teemoittelun avulla voidaan tutkia, ovatko jonkin kategorian sanat usein yhteydessä sijoittajien reaktioon lyhyellä aikavälillä.

5 TULOKSET

5.1 Regressioanalyysi

Seuraavaksi käydään läpi regressioanalyysin tuloksia koko aineiston osalta. Analyysi on suoritettu R-ohjelmiston versiolla 4.1.2. Taulukossa 5 on kuvattu regressiokertoimien arvoja, kun selitettävänä muuttujana on ollut epänormaali tuotto eri ajanjaksoilla. Sävytuottajista ainoastaan negatiivisuus saa tilastollisesti merkitsevästi nolasta eroavan kertoimen millään ajanjaksolla. Negatiivisuus vaikuttaisi olevan yhteydessä epänormaaleihin tuottoihin etenkin raportin julkaisupäivän (tapahtumapäivä 0) lähistöllä, ja yhteys näyttäisi laimenevan tarkasteluikkunan loppua kohden. Tapahtumapäivien 0 ja 1 osalta negatiivisuus saa kertoimen -1.262, mikä tarkoittaa, että jos muiden muuttujien ollessa vakioituja raportin negatiivisten sanojen osuus kasvaa yhden prosenttiyksikön, niin tämän ajanjakson epänormaali tuotto on keskimäärin 1.262 prosenttiyksikköä alhaisempaa. Kertoimen keskivirhe on noin 0.36, joten 95 % varmuudella epänormaalin tuoton lasku on 1.97 ja 0.56 prosenttiyksikön välillä, mitä voidaan pitää lyhyellä aikavälillä huomattavana vaikutuksena. Muilla ajanjaksoilla negatiivisuuden ja epänormaalien tuottojen yhteys on samankaltaista, lukuun ottamatta aikaväliä [5, 9], jolla negatiivisuuden kerroin ei enää eroa tilastollisesti merkitsevästi nolasta. Muista mallin 1 sävytuottajista ainoastaan epävarmuus aikavälillä [0, 1] oli lähellä tilastollista merkitsevyyttä ($p=0.08$), joten tämän tuloksen perusteella voisi olla pientä viitettä siihen, että epävarmuutta ilmaisevien sanojen osuuden kasvaessa myös epänormaalit tuotot kasvavat, mikä olisi aikaisempaan tutkimukseen nähden päinvastainen tulos, sillä epävarmuus on nähty riskisyyttä kuvaavana (ks. esim. Li, 2006), mutta toisaalta tulkinta epävarmuudesta mahdollisuuksien kuvaajana ei ole poissuljettua. Kuitenkin muilla tarkastelluilla aikaväleillä epävarmuuden kerroin ei ole yhtä lähellä tilastollista merkitsevyyttä, mikä vähentää syytä olettaa epävarmoiksi luokiteltujen sanojen osuuden olevan yhteydessä epänormaaleihin tuottoihin, sillä sen kerroin on mahdollisesti nolla. Sama tulkinta pätee myös positiivisille sanoille kaikilla tarkasteluun valituilla aikaväleillä.

Taulukko 5 Regressiomallin 1 selittävien muuttujien kertoimet ja mallin selitysaste, kun selitettävänä muuttujana on kumulatiivinen epänormaali tuotto eri ajanjaksoilla. Merkitsevyystasot: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Otoskoko 1118.

Muuttuja	Ajanjakso				
	[-10, 9]	[0, 1]	[0, 4]	[0, 9]	[5, 9]
(Vakio)	1.627	1.263	0.860	0.151	- 0.709
Positiivisuus	0.557	- 0.331	- 0.303	- 0.161	0.142
Negatiivisuus	- 2.673***	- 1.262***	- 0.940*	- 1.081*	- 0.141
Epävarmuus	1.220	1.075	0.742	0.813	0.072
EPS/hinta	0.149	0.115	0.085	0.041	- 0.045
Current Ratio muutos	- 0.305	- 0.177	- 0.168	- 0.030	0.138
Liikevoitto-% muutos	0.031***	0.015**	0.019**	0.024***	0.005
Velka/OPO muutos	- 0.016**	- 0.005	- 0.010**	0.014**	-0.004
Korjattu selitysaste	3.7 %	1.9 %	1.5 %	1.8 %	0.0 %

Painotettuja osuuksia (taulukko 6) tarkastellessa negatiivisuus on jälleen tilastollisesti merkitsevällä tasolla, nyt kuitenkin vain kahdella tarkastelujaksolla. Lisäksi painotettujen osuuksien kertoimet mallinnuksen tuloksena ovat mallin 1 muuttujiin nähden pienempiä, eli yhden yksikön muutos painotetun osuuden kertoimessa vastaa pienempää muutosta epänormaaleissa tuotoissa. Tätä voi selittää osin painotettujen arvojen suurempi vaihteluväli. Koko tarkastelujaksolla raportin painotetun negatiivisen arvon yhden yksikön kasvu on yhteydessä noin 0.209 prosenttiyksikön epänormaalien tuottojen laskuun, kun taas tapahtumapäivinä 0 ja 1 syntynyt epänormaali tuotto laskee noin 0.127 prosenttiyksikköä. Taulukoimattomia standardoituja kertoimia tarkastellessa huomataan kuitenkin, että aikavälillä [0, 1] painotettujen arvojen ja prosentti-osuuksien vaikutus on samaa suuruusluokkaa, sillä yhden keskihajonnan muutos vastaa 0.72 prosenttiyksikön laskua epänormaaleissa tuotoissa mallin 2 tapauksessa ja 0.68 laskua mallin 1 tapauksessa. Mallista 1 poiketen painotettu epävarmuus on tilastollisesti merkitsevä aikavälillä [0, 1], ja painotetun epävarmuuden kasvaessa yhden yksikön myös epänormaalit tuotot vaikuttaisivat nousevan 0.326 prosenttiyksikköä. Muilla tarkastelujaksoilla epävarmuuden kerroin ei ole tilastollisesti merkitsevä. Painotettu positiivisuus puolestaan ei yllä tilastolliseen merkitsevyyteen ainoallakaan tarkastelujaksolla.

Taulukko 6 Regressiomallin 2 selittävien muuttujien kertoimet ja mallin selitysaste, kun selitettävänä muuttujana on kumulatiivinen epänormaali tuotto eri ajanjaksoilla. Merkitsevyystasot: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Otokoko 1118.

Muuttuja	Ajanjakso				
	[-10, 9]	[0, 1]	[0, 4]	[0, 9]	[5, 9]
(Vakio)	0.370	0.399	0.574	- 0.308	- 0.882**
Pos_painotettu	0.243	- 0.024	- 0.081	- 0.045	0.036
Neg_painotettu	- 0.209**	- 0.127**	- 0.076	- 0.063	- 0.013
Epäv_painotettu	0.168	0.326*	0.129	0.115	- 0.014
EPS/hinta	0.150	0.113	0.081	0.041	- 0.040
Current Ratio muutos	- 0.299	- 0.179	- 0.161	- 0.026	0.135
Liikevoitto-% muutos	0.031***	0.015**	0.019**	0.024***	0.005
Velka/OPO muutos	- 0.016**	- 0.005	- 0.010**	- 0.014**	-0.004
Korjattu selitysaste	2.5 %	1.3 %	1.3 %	1.5 %	0.0 %

Taulukossa 7 on esitetty mallin 3 tuloksia. Negatiivisuuden muutoksen kerroin on tässäkin tapauksessa vahvasti tilastollisesti merkitsevä monella tarkasteluvälillä. Poikkeuksena muista malleista nyt myös positiivisuuden muutos on tilastollisesti merkitsevällä tasolla, ja sen vaikutus näyttäisi tapahtuvan hiukan jälkikäteen tapahtumapäivien 5 ja 9 välillä. Mikäli positiivisuuden muutos on yhden prosenttiyksikön suurempi kuin edellisellä tarkastelujaksolla, myös epänormaalit tuotot ovat keskimäärin noin 0.94 prosenttiyksikköä suuremmat, kun muut muuttujat ovat vakioituja. Kertoimen 95 % luottamusväli sijoittuu 0.21 ja 1.67 prosenttiyksikön välille. Negatiivisuuden muutoksen osalta suurin tilastollinen merkitsevyys on aikavälillä [0, 1], jolloin epänormaalien tuottojen lasku on 95 % varmuudella välillä -3.28 ja -1.04. Epävarmuuden muutoksella ei näyttäisi olevan vaikutusta epänormaalisiin tuottoihin millään ajanjaksolla.

Mallien selitysasteita vertailtaessa sävy muuttujien muutoksista koostuva malli 3 saa tapahtumapäivän jälkeisinä ajanjaksoina suurimmat korjatut selitysasteet. Kokonaisuutena mallien korjatut selitysasteet ovat melko alhaisia, pääosin yhden ja kahden prosentin välillä, mutta suurusluokka on kuitenkin samankaltainen aikaisemman tutkimuksen kanssa (ks. esim. Loughran & McDonald, 2011; Feldman ym., 2010).

Taulukko 7 Regressiomallin 3 selittävien muuttujien kertoimet ja mallin selitysaste, kun selitettävänä muuttujana on kumulatiivinen epänormaali tuotto eri ajanjaksoilla. Merkitsevyytasot: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Otokoko 1118.

Muuttuja	Ajanjakso				
	[-10, 9]	[0, 1]	[0, 4]	[0, 9]	[5, 9]
(Vakio)	0.163	0.120	- 0.119	- 0.732**	- 0.614***
Pos_muutos	0.958	- 0.561	- 0.913	0.027	0.940*
Neg_muutos	- 2.636**	- 2.162***	- 1.835**	- 1.896*	- 0.061
Epäv_muutos	- 0.299	0.219	0.343	- 0.207	- 0.550
EPS/hinta	0.158	0.117	0.086	0.042	- 0.043
Current Ratio muutos	- 0.352	- 0.207	- 0.189	- 0.059	0.129
Liikevoitto-% muutos	0.031***	0.015**	0.019**	0.023***	0.005
Velka/OPO muutos	- 0.016**	- 0.004	- 0.010*	0.013**	-0.004
Korjattu selitysaste	2.8 %	1.9 %	1.7 %	1.9 %	0.5 %

Seuraavaksi kuvataan regressiomallin 1 tuloksia, kun aineistoa tarkastellaan kahdessa osassa (taulukko 8). Aineiston alkuosa koostuu raporteista vuoden 2016 viimeisestä neljänneksestä vuoden 2018 viimeiseen neljännekseen, ja loppuosassa ovat raportit vuoden 2019 ensimmäisestä neljänneksestä vuoden 2020 viimeiseen neljännekseen. Kenties suurin ero alku- ja loppuosan aineistossa on tapahtumapäivien 0 ja 1 epänormaalien tuottojen ja epävarmuuden välinen yhteys. Vuoden 2019 jälkeen epävarmuuden kasvaminen raportissa on tilastollisesti merkitsevästi yhteydessä epänormaalien tuottojen kasvuun, mutta aineiston alkuosassa vastaavaa tilannetta ei ole. On mahdollista, että poikkeuksellisen koronapandemian aikana sijoittajat ovat ottaneet myönteisesti vastaan epävarmuudesta tai riskeistä viestimisen, mikäli toinen vaihtoehto olisi niiden sivuuttaminen raportoinnissa, kun taas nk. normaalioloissa epävarmuuksien esittäminen ei näyttäisi juurikaan vaikuttavan sijoittajien reaktioon. Toinen tilastollisesti merkitsevä sävymuuttujan kerroin on koko aineiston tavoin negatiivisuudella aikavälillä [0, 1], ja se pätee sekä alku- että loppuosassa aineistoa. Positiivisuus ei ole tilastollisesti merkitsevällä tasolla millään tarkastelujaksolla ennen eikä jälkeen vuotta 2019.

Taulukko 8 Regressiomalli 1 eri ajanjaksoilla, kun aineisto on jaettu kahteen osaan. Alkuosa q4 2016 – q4 2018 (N = 536), loppuosa q1 2019 – q4 2020 (N = 582). Merkitsevyytasot: *** p < 0.001, ** p < 0.01, * p < 0.05.

Ajanjakso Muuttuja	[0, 1]		[0, 4]		[0, 9]	
	Alku	Loppu	Alku	Loppu	Alku	Loppu
(Vakio)	1.974	0.532	0.815	0.647	0.453	-1.030
Positiivisuus	-0.708	0.066	-0.470	-0.165	-0.483	0.203
Negatiivisuus	-1.181*	-1.456**	-0.751	-1.180	-1.233	-0.864
Epävarmuus	-0.143	1.850*	0.107	1.104	0.818	0.756
EPS/hinta	0.016	0.161	0.060	0.101	0.052	0.063
Current Ratio muutos	-0.806	0.322	-1.321	0.863	-1.433	1.139
Liikevoitto-% muutos	0.022*	0.011	0.027**	0.014	0.030*	0.020*
Velka/OPO muutos	-0.007	-0.008	-0.012**	-0.014**	-0.041***	-0.002
Korjattu selityaste	2.1%	2.3 %	1.8 %	1.8%	8.1 %	0.0 %

5.2 Lähisanat

Taulukossa 9 on esitelty positiivisten sanojen lähistöllä olleita sanoja, jotka on teemoittelemalla koottu ryhmiksi. Tässä tarkastelussa aineisto on saatu hiukan laajempaan muotoon kuin regressioanalyysia tehdessä, sillä otoskokoa rajoittaneet kontrollimuuttujien tiedot sekä muuttujien muutoksia varten tehdyt laskut eivät vaikuta tähän tutkimuksen osioon. Jokaisen sanan kohdalla aineisto on jaettu kahtia sen mukaan, onko sana esiintynyt raporteissa positiivisen sanan läheisyydessä. Mainintojen lukumääriin ei ole kiinnitetty huomiota, vaan yksikin osuma tekstissä lisää raportin mainittuja tarkastelevaan joukkoon. Toisessa joukossa ovat sellaiset raportit, joissa joko sanaa ei ole esiintynyt lainkaan, tai se on esiintynyt, mutta sen lähistöllä ei ole ollut yhtään positiiviseksi luokiteltua sanaa. Tuloksiin on kerätty ne sanat, joiden maininnan ja mainitsemattomuuden yhteydessä tapahtuneet keskimääräiset tuotot eroavat todennäköisesti toisistaan. Kattavampien tulosten saamiseksi merkitsevyytasoksi on valittu 90 %, eli p-arvo 0.1 ja sen alle. Lisäksi molemmissa ryhmissä on vaadittu vähintään 10 havaintoa, jotta keskiarvovertailu olisi mielekkäämpää. Ryhmien välinen t-testi on suoritettu Pythonilla ja ryhmien välisiä variansseja ei ole oletettu yhteneväisiksi.

Taulukko 9 Positiivisten sanojen läheisyydessä olevat sanat, jotka tekstissä mainittuna ovat yhteydessä erilaiseen aikavälin [0, 4] epänormaaliin tuottoon mainitsemattomuuteen verrattuna. Otokoko 1455.

<i>Kehitys / kasvu</i>	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
<i>GROWTH</i>	0.0476	1291	- 0.044	264	- 1.153
<i>GROWING</i>	0.0478	334	0.556	1221	- 0.451
<i>DEVELOPMENT</i>	0.0282	1156	0.054	399	- 1.059
<i>EXPANSION</i>	0.0187	196	0.929	1359	- 0.394
<i>ACQUIRE</i>	0.0247	51	2.132	1504	- 0.319
<i>ACQUISITIONS</i>	0.0867	381	0.342	1174	- 0.425
<i>EXPANDING</i>	0.0659	120	0.963	1435	- 0.335
<i>Laskentatoimi</i>	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
<i>PROFIT</i>	0.0248	875	0.176	680	- 0.778
<i>LOSSES</i>	0.0006	319	1.134	1236	- 0.582
<i>ROE</i>	0.0156	19	2.063	1536	- 0.264
<i>GOODWILL</i>	0.0913	28	2.794	1527	- 0.294
<i>PAYABLES</i>	0.0718	15	- 2.667	1540	- 0.212
<i>Toiminta</i>	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
<i>PRODUCTION</i>	0.0149	512	- 0.953	1043	0.129
<i>QUALITY</i>	0.0409	480	0.386	1075	- 0.514
<i>SERVICES</i>	0.0478	778	0.173	777	- 0.646
<i>APPLICATIONS</i>	0.0439	112	1.365	1443	- 0.360
<i>SCALABILITY</i>	0.0744	28	1.630	1527	- 0.270
<i>RETAILERS</i>	0.0718	27	1.486	1528	- 0.266
<i>MARKETS</i>	0.0436	649	0.248	906	- 0.591
<i>INFRASTRUCTURE</i>	0.0889	134	- 1.298	1421	- 0.133
<i>CUSTOMERS</i>	0.0617	895	0.101	660	- 0.688
<i>CUSTOMER</i>	0.0878	884	0.072	671	- 0.644
<i>PERFORMANCE</i>	0.0539	1015	0.060	540	- 0.805
<i>MACHINERY</i>	0.0109	59	- 1.891	1496	- 0.166
<i>COMMODITY</i>	0.0372	42	2.445	1513	- 0.314
<i>Ajanjakso</i>	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
<i>PERIOD</i>	0.0344	1149	0.019	406	- 0.956
<i>QUARTERLY</i>	0.0003	112	2.701	1443	- 0.466
<i>QUARTER</i>	0.0017	1218	0.129	337	- 1.523
<i>MONTHS</i>	0.0527	240	0.757	1315	- 0.419
<i>Näkymät</i>	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
<i>VIEW</i>	0.0025	123	- 2.066	1432	- 0.078
<i>FORECASTED</i>	0.0082	12	- 5.810	1543	- 0.193
<i>FORECASTS</i>	0.0977	60	- 1.698	1495	- 0.177

Ympäristö, aineeton pääoma	p-arvo	On mainittu		Ei mainintaa	
		N	Keskiarvo	N	Keskiarvo
ESG	0.0905	28	1.819	1527	- 0.274
RENEWABLE	0.0924	102	0.829	1453	- 0.315
REPUTATION	0.0801	93	1.070	1462	- 0.315
TALENT	0.0819	74	1.372	1481	- 0.318
EMISSIONS	0.0233	114	1.230	1441	- 0.355

Ensimmäinen yhdistetty ryhmä koostuu kasvua ja kehitystä kuvaavista sanoista. Näistä eniten positiivisten sanojen lähistöllä esiintyivät kasvu (*growth*) sekä kehitys (*development*), ja näin tapahtuessa epänormaalien tuottojen keskiarvo oli molemmissa tapauksissa lähellä nollaa. Päinvastaisessa tapauksessa epänormaalit tuotot olivat keskimäärin noin 1.2 prosenttiyksikköä negatiivisen puolella, joten vaikuttaisi siltä, että positiiviseen sävyyn kasvusta ja kehityksestä puhuttaessa sijoittajien reaktio on neutraali, mutta sen puuttuessa reaktio kääntyy negatiiviseksi. Myös muiden tämän ryhmän sanojen tapauksissa lähistöllä ollut positiivinen sana viittaisi korkeampiin epänormaaleihin tuottoihin, ja sijoittajien reaktio vaikuttaisi olevan huomattavan myönteinen etenkin laajentumisesta (*expansion*) tai hankkimisesta (*acquire*) raportoitaessa.

Toinen havaittujen sanojen ryhmä sisältää laskentatoimeen liittyviä käsitteitä. Liikevoiton (*profit*) osalta on huomattavissa samankaltainen ilmiö kuin kasvun tapauksessa, sillä positiivinen maininta siitä ei keskimäärin johda kovin suuresti positiivisiin epänormaaleihin tuottoihin, mutta sen puuttuminen vaikuttaisi olevan yhteydessä keskimäärin noin 0.78 prosenttiyksikön odotettua alempiin tuottoihin. Positiivinen maininta oman pääoman tuoton (*ROE*) sekä liikearvon (*goodwill*) lähistöllä sen sijaan on yhteydessä peräti yli kahden prosenttiyksikön odotettua suurempiin tuottoihin, joskin havaintojen määrä näissä kahdessa tapauksessa ei ole järin suuri. Sen sijaan positiivinen sana ostovelkojen (*[accounts] payables*) lähistöllä on yhteydessä suurehkoihin negatiivisiin epänormaaleihin tuottoihin, vaikkakin jälleen havaintojen määrä on pieni ja tulos ei aivan ole tilastollisesti merkitsevä. Vaikeammin tulkittavissa on tappioiden (*losses*) tapaus, sillä positiivisen sanan esiintyessä sen lähistöllä epänormaalit tuotot vaikuttaisivat positiivisilta, kun taas esiintymättömyyden tilanteessa tuotot ovat negatiivisen puolella, mikä on esimerkiksi kasvaneista tappioista puhuttaessa hiukan intuition vastaista. On mahdollista, että tappioihin on tekstissä yhdistetty syitä, joiden ilmaisemiseen sijoittajat ovat reagoineet myönteisesti, mutta toisaalta muut lähistöllä olleet sanat ovat saattaneet tässä tapauksessa vaikuttaa tuloksiin.

Toimintaa kuvailevia sanoja esiintyi tuloksissa muita ryhmiä enemmän. Aiempien ryhmien tavoin useasti positiivisen sanan lähistöllä mainitut sanat, kuten asiakkaat ja suorituskyky (*customer(s), performance*), eivät keskimäärin ole yhteydessä erityisen suurin epänormaaleihin tuottoihin, mutta maininnan puuttuessa epänormaalit tuotot ovat keskiarvoisesti negatiivisia. Muutamit harvemmin positiivisen sanan lähistöllä esiintyneet sanat vaikuttaisivat esiintyessään olevan yhteydessä odotettua suurempiin tuottoihin, esimerkkeinä skaa-

lautuvuus (*scalability*) ja jälleenmyyjät (*retailers*). Toisaalta myös vastaesimerkkejä löytyy, kuten koneiden (*machinery*) tapauksessa. Toimintaa koskevat sanat saattavat riippua jonkin verran yrityksen toimialasta, mikä hiukan rajoittaa yleisemmällä tasolla tehtäviä päätelmiä tämän sanaryhmän ja siihen yhdistettävissä olevien epänormaalien tuottojen tulkintaa.

Osa havainnoiksi tulleista sanoista kuvaa puhtaasti osavuosikatsauksen tarkastelujaksoa, jonka lähistöllä on aineiston perusteella ollut usein jokin positiiviseksi luokiteltu sana, kuten neljänneksen ja jakson (*quarter, period*) tapauksissa. Suuri havaintojen määrä positiivisen sanan lähellä ei näyttäisi johtavan nollasta eroaviin epänormaaleihin tuottoihin, mutta tämän yhdistelmän puuttuminen voitaisiin yhdistää odotettua matalampiin tuottoihin. Havainnoissa oli myös muutama näkymiä koskeva sana, mutta mainintojen määrä on melko vähäinen, mikä osin vaikeuttaa kohtuullisen suuresti negatiivisten keskimääräisten epänormaalien tuottojen tulkintaa. Lisäksi yksi tämän ryhmän sanoista (*view*) on melko monitulkintainen, joten maininta ei välttämättä ole viitannut esimerkiksi paraneviin näkymiin. Viimeisenä ryhmänä havainnoista eriytyi ympäristöön ja aineettomaan pääomaan viittaavia sanoja. Tässä ryhmässä mainintojen määrät positiivisen sanan lähistöllä ovat suhteellisen vähäisiä, mutta näissä tapauksissa maininnat ovat yhdistettävissä positiivisiin epänormaaleihin tuottoihin. Muutoin yhteydet sanojen ja tuottojen välillä ovat helpohkosti ymmärrettävissä, mutta päästöjen (*emissions*) yhteys odotettua suurempiin tuottoihin on hiukan kyseenalaista, esimerkiksi jos tekstissä olisi mainittu kasvaneet päästöt.

Tämän tarkastelun tuloksena syntynyt sanalista lienee ennemminkin vain suuntaa antava, ja yksittäisten sanojen ja niiden yhteydessä tapahtuneiden epänormaalien tuottojen tulkinnassa on epävarmuutta. Lähisanoja tallentaessa jokaisen sanan vieressä on myös muita kuin positiiviseksi tunnistettuja sanoja, jotka voivat muuttaa kahden havaitun sanan yhteyttä. Lisäksi tämänkaltainen tarkastelu ei ota huomioon mahdollisia negaatioita sanojen vieressä, jotka voisivat kääntää maininnan merkityksen pääläelleen. Toisaalta kuitenkin havainnoissa ilmeni muutamia sellaisia sanoja, joista oli useita mainintoja, jolloin väärin tulkitsemisen riski pienenee. Näissä tapauksissa keskimääräiset epänormaalit tuotot olivat usein lähellä nollaa, kun taas maininnan puuttuessa negatiivisia, mikä viittäisi siihen, että maininta itsessään ei välttämättä ole sijoittajan kannalta merkityksellinen, mutta joko lähistöllä oleva positiivisuuden tai koko sanan puute saattaa implikaatioillaan vähentää sijoittajien arvostusta yritystä kohtaan.

6 JOHTOPÄÄTÖKSET

Tässä tutkimuksessa on tarkasteltu suomalaisten yritysten osavuosisikastausten julkaisuhetkillä tapahtuneiden epänormaalien tuottojen sekä katsausten tekstuaalisen sävyn välistä yhteyttä. Sävy on määritelty laskemalla positiiviseksi, negatiiviseksi ja epävarmuutta kuvaaviksi luokiteltuja sanoja raporttiaineistosta. Yhteyttä tutkittiin kahden tutkimuskysymyksen avulla:

- Onko suomalaisten talousraporttien sävyllä ja niiden julkaisuhetkellä muodostuvilla lyhyen aikavälin epänormaaleilla osaketuotoilla yhteyttä?
- Mitkä sanat tai sanaryhmät positiivisten sanojen lähistöllä ovat yhteydessä poikkeavaan reaktioon osakemarkkinoilla?

Ensimmäistä tutkimuskysymystä tarkasteltiin regressiomallien avulla, joissa kolmen eri tyyppin sävyymuuttujia liitettiin yhteen yrityksen taloudellista tilaa kuvaavien kontrollimuuttujien kanssa, jotta raporttien sävyn vaikutusta oli mahdollista tarkastella huolimatta yrityksen taloudellisista muutoksista. Ensimmäisessä mallissa sävyymuuttujat määrittyivät prosenttiosuuksien mukaan, toisessa koko aineistossa tapahtuneen esiintymistiheyden mukaan ja kolmannessa mallissa prosenttiosuuden muutoksen mukaan. Kaikissa malleissa selitettävänä muuttujana oli raportin julkaisuhetkellä tapahtunut epänormaali tuotto eri ajanhetkinä julkaisun lähistöllä. Otokokona regressiomallinnuksessa oli 1118 havaintoa.

Kaikkien kolmen regressiomallin tulosten perusteella raportoinnin negatiivisuus ja sen muutos vaikuttaisi olevan yhteydessä aleneviin raportin julkaisuhetken epänormaaleihin tuottoihin, ja sijoittajat ainakin näyttäisivät osakekurssin muutosten perusteella ottavan huomioon raportoinnin negatiivisuuden päätöksenteossaan. Tulos on aikaisemman tutkimuksen kaltainen, sillä etenkin negatiivisuus on noussut esiin osakekurssien muutosten selittäjänä ja yhteys on ollut juuri epänormaaleja tuottoja laskeva (ks. esim Loughran & McDonald, 2011). Epävarmuuden vähäinen merkitys poikkeaa hieman aikaisemmista havainnoista, sillä Li (2006) havaitsi riskin ja epävarmuuden pystyvän ennustamaan osaketuottoja, mutta toisaalta tämä tarkastelu ylsi pidemmälle aikavälille. Toisaalta, kun aineistoa tarkasteltiin vuoden 2019 alusta, eli koronapandemian ajalta, epävarmoiksi luokiteltujen sanojen osuus raporteissa oli tilastollisesti

merkitsevällä tasolla positiivisesti yhteydessä epänormaalien osaketuottojen kanssa. Tämä havainto eroaa Lin (2006) tuloksista, joissa epävarmuuden ja osaketuottojen yhteys oli negatiivinen. Eroavaisuus saattaa selittyä poikkeusolojen aiheuttamalla yleisellä epätietoisuudella, jossa oman toimintansa epävarmuuksia tunnistaneet ja niistä viestineet yritykset saatiin sijoittajien toimesta nähdä turvallisimpina sijoituskohteina. Negatiivisen sävyn ja epänormaalien tuottojen yhteys oli tilastollisesti merkitsevällä tasolla sekä ennen että jälkeen koronapandemian. Positiivinen sävy raportoinnissa ei sen sijaan osoittautunut tämän tutkimuksen perusteella juurikaan merkitykselliseksi sijoittajien päätöksenteon kannalta, kun taas aikaisemmassa tutkimuksessa näyttöä on ollut sekä positiivisuuden ja osaketuottojen yhteyden puolesta (Jegadeesh & Wu, 2012) että vastaan (Loughran & McDonald, 2011), mikä voi osin johtua positiivisten sanojen vaikeasta luokittelusta (Loughran & McDonald, 2016). Näistä tuloksista voidaan päätellä, että sijoittajalla on mahdollisuus lyhyen aikavälin ylituottoihin sävyinformaatiota, etenkin negatiivisuuden suhteen, tarkastelemalla.

Vaikka aiemmassa laskentatoimen tekstianalyysin tutkimuksessa on arvioitu tekstimuotoisen informaation olevan hitaasti osakkeen hintoihin siirtyvää (ks. esim. Cohen ym., 2020; Li, 2006), negatiivisuuden ja epänormaalien tuottojen yhteys on tässä tutkimuksessa havaittu voimakkaimmaksi juuri raportin julkaisu- ja sitä seuraavana päivänä. Tämäkään ei ole täysin uusi havainto, vaan myös Feldmanin ym. (2010) tuloksissa lyhyen aikavälin epänormaali tuotot ja raportoinnin sävy ovat kytköksissä. On mahdollista, että negatiivinen sävy ja etenkin sen muutokset raportoinnin tekstiosuuksissa huomataan ja tulkitaan nopeasti osakemarkkinoilla, mutta kaikki tekstimuotoisen informaation ominaisuudet, kuten Cohenin ym. (2020) tarkastelemat muutokset, eivät ole yhtä nopeasti havaittavissa.

Toista tutkimuskysymystä tarkasteltiin jakamalla aineisto jokaisen positiiviseksi tunnistetun sanan lähistöllä olleen sanan kohdalla kahtia sen mukaisesti, onko kyseinen yhdistelmä esiintynyt raportissa vai ei. Näin syntyneiden kahden raporttiryhmän välisten keskimääräisten epänormaalien tuottojen eroavaisuutta tutkittiin kahden otoksen t-testin avulla, jolloin tuloksena saatiin lista sanoista, jotka ovat yhteydessä eroavaan epänormaaliin tuottoon, mikäli niiden lähellä on raportin tekstissä esiintynyt jokin positiiviseksi luokiteltu sana. Tästä listasta koostettiin sanaryhmiä teemoittain. Otokoko tässä tarkastelussa oli hiukan suurempi kuin regressioanalyysien tapauksessa, yhteensä 1555 eri raporttia.

Tulokseksi tulleesta sanalistasta erottautui muutamia ryhmiä. Operaatioonaliseen toimintaan liittyvä sanaryhmä oli suurin näistä, mikä on linjassa Lin (2010) havaintojen kanssa, joiden mukaan prosenttiosuudeltaan suurin yksittäinen raporttien tekstiosuuksissa esiintyvä kategoria on yrityksen toimintaa kuvaava, joskin tämä tarkastelu tehtiin lausetasolla. Barthin ym. (2021) mukaan talousraporttien arvorelevanssiin vaikuttaa olennaisesti perinteisesti tarkastelujen liikevoiton ja kassavirtojen lisäksi aineeton pääoma ja kasvunäkymät, mikä oli huomattavissa myös tämän lähisanatarkastelun tuloksissa. Keskimääräiset epänormaali tuotot erosivat tilastollisesti merkitsevällä tavalla muutamien

kasvua ja laajentumista kuvaavien sanojen tapauksissa, lisäksi aineettomaksi pääomaksi luokiteltavia sanoja oli päätyynyt tuloksiin. Yleistä laskentajärjestelmää on kritisoitu sen sopimattomuudesta kuvaamaan ja mittaamaan aineettoman pääoman merkitystä (Lev, 2019), joten on mahdollista, että yritykset pyrkivät raportoinnin tekstiosuuksissa laajentamaan numeroiden antamaa kuvaa esimerkiksi omista kasvumahdollisuuksistaan, jonka taas tämän tutkimuksen perusteella sijoittajat myös noteeraavat.

Toinen olennaisista lähisanatarkastelun havainnoista on se, että joidenkin sanojen läheisyydessä positiivinen maininta on niin yleinen, että sen yhteydessä tapahtuvien epänormaalien tuottojen suuntaa ei voi päätellä, toisin sanoen ne ovat keskimäärin lähellä nollaa. Sen sijaan, mikäli sanaa ei ole mainittu positiivisen sanan lähistöllä raportoinnissa, epänormaali tuotto näyttäisivät olevan alemmalla tasolla. Esimerkkejä tällaisista sanoista olivat tulosten perusteella liikevoitto, neljännes ja kasvu. Toisaalta tuloksissa esiintyi myös sellaisia sanoja, jotka positiivisen sanan läheisyydessä esiintyessään olivat yhteydessä vahvasti nollostä eroaviin epänormaaleihin tuottoihin. Näistä sanoista oli tyypillisesti vähän havaintoja aineistossa, esimerkkeinä oman pääoman tuotto, skaalautuvuus ja hankkia.

Tähän tutkimukseen liittyy rajoitteita, jotka tulee ottaa huomioon tuloksien merkitystä pohdittaessa. Tutkimuksessa käytetty tekstianalyysin menetelmä, eli luokiteltujen sanojen esiintymistiheyksien laskeminen, jättää huomiotta sanojen väliset suhteet ja niiden merkitykset, mikä voi johtaa joissakin tapauksissa sävyn virheelliseen luokitteluun. Toisaalta läpikäytyjen sanojen määrä on erittäin suuri ja niiden kategorisoimiseen on käytetty aikaisemmassa tutkimuksessa testattua luokittelulistaa, mikä vähentää riskiä sävyn mitattavuuden epätarkkuudesta. Kuitenkaan, vaikka jokainen sana saataisiin luokiteltua täsmällisesti, ei voida olla varmoja, millä tavalla juuri luokiteltujen sanojen määrä vaikuttaa osaketuottoihin, sillä regressioanalyysin kertoimista ei voi päätellä kausaalisuuden suuntaa, vaikka muuttujien välinen yhteys ilmenisikin. Huomattavaa on myös, että epänormaaleihin tuottoihin vaikuttavat monet seikat, ja vaikka niitä on tässä tutkimuksessa kontrollimuuttujien avulla pyritty kartoittamaan, monta tuntematonta muuttujaa saattaa jäädä regressiomallin ulkopuolelle, mikä voi osin näkyä tämän tutkimuksen mallien alhaisissa selityksasteissa. Lisäksi tämän tutkimuksen tarkastelu olettaa, että sijoittajat pystyvät käsittelemään tarkasti ja rationaalisesti käytännössä jokaista raportoinnissa esiintyvää sanaa tai lausetta, ja samaan aikaan nämä sanat olisivat yhtä kuvaavia raportin sävyn kannalta, mitä aikaisemman tutkimuksen perusteella voi olla haastavaa perustella (ks. esim. Hirschleifer & Teoh, 2003). Toisaalta raportoinnin sävyn voi nähdä myös kokonaisvaltaisempana käsitteenä, jossa yksittäisen sanan esiintyminen tai puuttuminen ei liikauta sijoittajan arviota suuresti, vaan se muodostuu sävyä kuvaavien sanojen poikkeuksellisuudesta tai määrällisestä ylivoimaisuudesta. Tällöin sijoittaja saattaa olla altis vaikutelmien hallinnoinnille, eli yrityksen tietoiselle tekstimuotoisen informaation kaunistelulle, jonka seurauksista on vaihtelevaa näyttöä aikaisemmassa tutkimuksessa (ks. esim. Merkl-Davies & Brennan, 2007).

Myöskään tutkimuksessa suoritettu lähisanatarkastelu ei ole täysin aukoton, sillä kahden lähekkäisen sanan merkitykseen vaikuttavat luonnollisesti muut lähistöllä olevat sanat, jotka saattavat muuttaa tavoiteltua positiivisuuden ja tietyn sanan välistä sanomaa. Toisaalta menetelmänä käytetyn kahden otoksen t-testin avulla voidaan todeta vain keskimääräisten epänormaalien tuottojen ero, mutta muita epänormaaleihin tuottoihin vaikuttavia tekijöitä ei voida ottaa huomioon. Lisäksi haastetta yksittäisen sanan lähistöllä esiintyvän positiivisuuden tulkintaan aiheuttaa itse sanan merkitys tai tarkoitus, varsinkin kun tässä tutkimuksessa käytettiin suomalaisten yhtiöiden englannin kielellä tehtyjä raportteja.

Tutkimuksen rajoitteet antavat aiheita jatkotutkimukselle, joista yhtenä nousee esiin raportoinnin sävyn määrittämisen kehittäminen. Vaikka sanalis-toilla, etenkin talousmaailman sopivilla, on saatu lupaavia tuloksia sävyn arvorelevanssista, kuvaavampia tuloksia saatetaan saada ottamalla enemmän huomioon sanojen väliset suhteet ja lausetason merkitykset, kenties teko-älyavusteisesti. Lähisanatarkastelua voi olla mahdollista kehittää esimerkiksi ottamalla huomioon mainintojen määrä raportissa tai positiivisen ja tarkastelun sanan välinen etäisyys. Lisäksi tarkastelua voi olla hyödyllistä kohdistaa tässäkin tutkimuksessa esiin nousseisiin, vaikeasti mitattavia määreitä sisältäviin aihealueisiin, kuten kasvunäkymät tai aineeton pääoma. Tekstianalyysin tutkimuksen kannalta olennaista on informaation keräämisen ja päätöksenteon välisen mekanismin, jota kvalitatiivisen informaation tapauksessa voidaan pitää monimutkaisena, kattavampi ymmärrys, jotta voitaisiin päätellä, missä määrin tekstimuodossa annettu tieto on käsitettävissä ja kuinka suuri rooli sillä on sijoittajien päätöksenteon kannalta. Jatkotutkimusta on mahdollista tehdä myös tekstianalyysia laajentamalla muihin konteksteihin suomalaisten yritysten osalta, esimerkiksi yrityksen suorituskyvyn ja raportoinnin sävyn yhteys pidemällä aikavälillä voisi olla mielenkiintoinen tarkastelun kohde. Toisaalta raporteista löytyy lukujen ja tekstin ohella muutakin sijoittajan päätöksentekoon mahdollisesti vaikuttavaa materiaalia, kuten kuvioita, taulukoita tai muuta graafista aineistoa, jota voisi olla mahdollista tutkia esimerkiksi informaation välittymisen tai sijoittajien tarkkaavaisuuden näkökulmasta.

LÄHTEET

- Abrahamson, E., & Amir, E. 1996. The information content of the president's letter to shareholders. *Journal of Business Finance and Accounting*, 23, 1157-1182.
- Aerts, W. 1994. On the use of accounting logic as an explanatory category in narrative accounting disclosures. *Accounting, organizations and society*, 19(4-5), 337-353.
- Ahmad, K., Han, J., Hutson, E., Kearney, C., & Liu, S. 2016. Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, 37, 152-172.
- Allee, K. D., & DeAngelis, M. D. 2015. The structure of voluntary disclosure narratives: Evidence from tone dispersion. *Journal of Accounting Research*, 53(2), 241-274.
- Alves, C. F. & Teixeira Dos Santos, F. 2008. Do first and third quarter unaudited financial reports matter? the portuguese case. *The European Accounting Review*, 17(2), 361-392.
- Amir, E., & Lev, B. 1996. Value-relevance of nonfinancial information: The wireless communications industry. *Journal of accounting and economics*, 22(1-3), 3-30.
- Baginski, S. P., Hassell, J. M., & Hillison, W. A. 2000. Voluntary causal disclosures: Tendencies and capital market reaction. *Review of quantitative finance and accounting*, 15(4), 371-389.
- Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. 2004. Why do managers explain their earnings forecasts?. *Journal of accounting research*, 42(1), 1-29.
- Ball, R. & Brown, P. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2), 159-178.
- Baird, J.E. and R.C. Zelin. 2000. The effects of information ordering on investor perceptions: An experiment utilizing presidents' letters. *Journal of Financial and Strategic Decisions* 13 (3): 71-81.
- Barth, M. E., Beaver, W. H., & Landsman, W. R. 2001. The relevance of the value relevance literature for financial accounting standard setting: another view. *Journal of accounting and economics*, 31(1-3), 77-104.
- Barth, M. E., Li, K., & McClure, C. 2021. Evolution in value relevance of accounting information (Working paper).
- Beaver, W., McNichols, M. & Wang, Z. 2018. The information content of earnings announcements: New insights from intertemporal and cross-sectional behavior. *Review of Accounting Studies*, 23(1), 95-135.
- Benninga, S. 2014. *Financial modeling* (4). Cambridge, Massachusetts: The MIT Press.
- Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. 2010. The financial reporting environment: Review of the recent literature. *Journal of accounting and economics*, 50(2-3), 296-343.

- Bochkay, K. & Levine, C. B. 2019. Using MD&A to improve earnings forecasts. *Journal of Accounting, Auditing & Finance*, 34(3), 458-482.
- Bozanic, Z., Roulstone, D. T. & Van Buskirk, A. 2018. Management earnings forecasts and other forward-looking statements. *Journal of Accounting & Economics*, 65(1), 1-20.
- Boudt, K., & Thewissen, J. 2019. Jockeying for position in CEO letters: Impression management and sentiment analytics. *Financial Management*, 48(1), 77-115.
- Brown, L. D. & Rozeff, M. S. 1979. The predictive value of interim reports for improving forecasts of future quarterly earnings. *The Accounting Review*, 54(3), 585-591.
- Brown, S. V. & Tucker, J. W. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2), 309-346.
- Bushee, B. J., Gow, I. D. & Taylor, D. J. 2018. Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, 56(1), 85-121.
- Butler, M., Kraft, A. & Weiss, I. S. 2007. The effect of reporting frequency on the timeliness of earnings: The cases of voluntary and mandatory interim reports. *Journal of Accounting & Economics*, 43(2), 181-217.
- Cheng, L., Roulstone, D. T., & Van Buskirk, A. 2021. Are Investors Influenced by the Order of Information in Earnings Press Releases?. *The Accounting Review*, 96(2), 413-433.
- Clatworthy, M., & Jones, M. J. 2003. Financial reporting of good news and bad news: evidence from accounting narratives. *Accounting and business research*, 33(3), 171-185.
- Cohen, L., Malloy, C. & Nguyen, Q. 2020. Lazy prices. *The Journal of Finance (New York)*, 75(3), 1371-1415.
- Davis, A. K., Piger, J. M. & Sedor, L. M. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 29(3), 845-868.
- Davis, A. K., & Tama - Sweet, I. 2012. Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, 29(3), 804-837.
- Diamond, D. W. 1985. Optimal release of information by firms. *The Journal of Finance (New York)*, 40(4), 1071-1094.
- Durnev, A. & Mangen, C. 2011. The real effects of disclosure tone: Evidence from restatements. *SSRN Electronic Journal*, doi:10.2139/ssrn.1650003
- Elliott, W.B. 2006. Are investors influenced by pro forma emphasis and reconciliations in earnings announcements? *The Accounting Review* 81 (1): 113-133.
- Eskola, J. & Suoranta, J. 1998. Johdatus laadulliseen tutkimukseen. *Vastapaino*.
- Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance (New York)*, 25(2), 383-417.
- Fama, E. F. & French, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.

- Feldman, R., Govindaraj, S., Livnat, J. & Segal, B. 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15(4), 915-953.
- Francis, J., Nanda, D. & Olsson, P. 2008. Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research*, 46(1), 53-99.
- Francis, J., & Schipper, K. 1999. Have Financial Statements Lost Their Relevance? *Journal of Accounting Research (Wiley-Blackwell)*, 37(2), 319-352.
- Francis, J., Schipper, K., & Vincent, L. 2002. Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review*, 77(3), 515-546.
- French, K. 2021. Developed Markets Factors and Returns. [sähköinen tutkimusaineisto]. Haettu osoitteesta http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed
- Fu, R., Kraft, A. & Zhang, H. 2012. Financial reporting frequency, information asymmetry, and the cost of equity. *Journal of Accounting & Economics*, 54(2-3), 132-149.
- Griffin, J. M. 2002. Are the Fama and French factors global or country specific?. *The Review of Financial Studies*, 15(3), 783-803.
- Griffin, P. 2003. Got information? investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies*, 8(4), 433-460.
- Grossman, S. J., & Stiglitz, J. E. 1976. Information and competitive price systems. *The American Economic Review*, 66(2), 246-253.
- Haaramo, V., Palmuaro, S. & Peill, E. 2018. Pörssiyrityksen tilinpäätös. Helsinki: Alma Talent.
- Hadro, D., Klimczak, K. M., & Pauka, M. 2017. Impression Management in Letters to Shareholders: Evidence from Poland. *Accounting in Europe*, 14(3), 305-330.
- Hajek, P., Olej, V. & Myskova, R. 2014. Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy*, 20(4), 721-738.
- Hand, J. R. (1990). A test of the extended functional fixation hypothesis. *Accounting Review*, 740-763.
- Healy, P. M. & Palepu, K. G. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting & Economics*, 31(1), 405-440.
- Hirshleifer, D., & Teoh, S. H. 2003. Limited attention, information disclosure, and financial reporting. *Journal of accounting and economics*, 36(1-3), 337-386.
- Holden, C. W., & Subrahmanyam, A. 1992. Long - lived private information and imperfect competition. *The Journal of Finance*, 47(1), 247-270.
- Huang, X., Teoh, S. H., & Zhang, Y. 2014. Tone management. *The Accounting Review*, 89(3), 1083-1113.

- Huang, A. H., Zang, A. Y., & Zheng, R. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review*, 89(6), 2151-2180.
- Hutton, A. P., Miller, G. S., & Skinner, D. J. 2003. The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research*, 41(5), 867-890.
- Jegadeesh, N. & Wu, D. 2013. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3), 712-729.
- Jiang, F., Lee, J., Martin, X. & Zhou, G. 2019. Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1), 126-149.
- Kaakinen M. & Ellonen, N. 2021. Regressioanalyysi. Teoksessa Kvantitatiivisen tutkimuksen verkkokäsikirja. Tampere: Yhteiskuntatieteellinen tietoaarkisto [ylläpitäjä ja tuottaja]. Haettu osoitteesta <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/analyysi/>
- Kajüter, P., Klassmann, F. & Nienhaus, M. 2016. Do reviews by external auditors improve the information content of interim financial statements? *The International Journal of Accounting Education and Research*, 51(1), 23-50.
- Kajüter, P., Lessenich, A., Nienhaus, M. & van Gemmern, F. 2021. Consequences of interim reporting: A Literature review and future research directions. *The European Accounting Review*, ahead-of-print(ahead-of-print), 1-31.
- Kearney, C. & Liu, S. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171-185.
- Kim, O. & Verrecchia, R. E. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting & Economics*, 17(1), 41-67.
- Lee, Y. 2012. The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research*, 29(4), 1137-1170.
- Leftwich, R. W., Watts, R. L. & Zimmerman, J. L. 1981. Voluntary corporate disclosure: The case of interim reporting. *Journal of Accounting Research*, 19(2), 50-77.
- Leung, S., Parker, L., & Curtis, J. 2015. Impression management through minimal narrative disclosure in annual reports. *The British accounting review*, 47(3), 275-289.
- Lev, B. 2019. Ending the accounting-for-intangibles status quo. *European Accounting Review*, 28(4), 713-736.
- Lev, B., & Zarowin, P. 1999. The boundaries of financial reporting and how to extend them. *Journal of Accounting research*, 37(2), 353-385.
- Li, E. X. & Ramesh, K. 2009. Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review*, 84(4), 1171-1208.
- Li, F. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? *SSRN Electronic Journal*,
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, 45(2-3), 221-247.

- Li, F. 2010. The information content of forward-looking statements in corporate filings-A naïve bayesian machine learning approach. *Journal of Accounting Research*, 48(5), 1049-1102.
- Li, J. & Zhao, X. 2016. Complexity and information content of financial disclosures: Evidence from evolution of uncertainty following 10-K filings. *SSRN Electronic Journal*,
- Li, M., Wu, H., Xiao, M. & You, J. 2019. Beyond cheap talk: Management's informative tone in corporate disclosures. *Accounting and Finance (Parkville)*, 59(5), 2905-2959.
- Lightstone, K., Young, N. M. & Mcfadden, T. 2012. Information quality of interim financial statements. *Accounting Perspectives*, 11(4), 297-313.
- Loughran, T. & McDonald, B. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance (New York)*, 66(1), 35-65.
- Loughran, T. & McDonald, B. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.
- Loughran, T. & McDonald, B. 2020. Loughran-McDonald Master Dictionary. [sähköinen tutkimusaineisto]. Haettu osoitteesta <https://sraf.nd.edu/textual-analysis/code/>
- MacKinlay, A. C. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature*, 35 (1), 13-39.
- Marty, T., Vanstone, B., & Hahn, T. 2020. News media analytics in finance: a survey. *Accounting & Finance*, 60(2), 1385-1434.
- Matsumoto, D., Pronk, M., & Roelofsen, E. 2011. What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, 86(4), 1383-1414.
- McEwen, R. A. & Schwartz, B. N. 1992. Are firms complying with the minimum standards for interim financial reporting? *Accounting Horizons*, 6(1), 75.
- Mendenhall, R. R. & Nichols, W. D. 1988. Bad news and differential market reactions to announcements of earlier-quarters versus fourth-quarter earnings. *Journal of Accounting Research*, 26(2), 63-86.
- Mensah, Y. & Werner, R. 2008. The capital market implications of the frequency of interim financial reporting: An international analysis. *Review of Quantitative Finance and Accounting*, 31(1), 71-104.
- Merkel-Davies, D., & Brennan, N. 2007. Discretionary disclosure strategies in corporate narratives: Incremental information or impression management. *Journal of Accounting Literature*, 26, 116-196.
- Merkley, K. J. 2014. Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review*, 89(2), 725-757.
- Milgrom, P. R. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380-391.
- Miller, B. P. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6), 2107-2143.

- Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. 2015. Forward-looking MD&A disclosures and the information environment. *Management Science*, 61(5), 931-948.
- Nasdaq Nordic. 2021a. Osakkeet. [sähköinen tutkimusaineisto]. Haettu osoitteesta <http://www.nasdaqomxnordic.com/osakkeet>
- Nasdaq Nordic. 2021b. Yhtiötiedotteet. [sähköinen tutkimusaineisto]. Haettu osoitteesta <http://www.nasdaqomxnordic.com/uutiset/yhtiötiedotteet>
- Parkkonen, J. & Knuts, M. 2014. Arvopaperimarkkinalaki (5., uud. p.). Helsinki: Talentum.
- Pastena, V. & Ronen, J. 1979. Some hypotheses on the pattern of management's informal disclosures. *Journal of Accounting Research*, 17(2), 550-564.
- Rogers, R. & Grant J. 1997. Content analysis of information cited in reports of sell-side financial analysts. *Journal of Financial Statement Analysis*, 3(1), 17.
- Rutherford, B. A. 2005. Genre analysis of corporate annual report narratives: A corpus linguistics-based approach. *The Journal of Business Communication*, 42(4), 349-378.
- Schleicher, T. & Walker, M. 2015. Are interim management statements redundant? *Accounting and Business Research*, 45(2), 229-255.
- Schloetzer, J. D., Tseng, A., Yohn, T. L., & Yoon, Y. S. 2021. Blame attribution and disclosure propensity. *The Accounting Review*, 96(4), 405-432.
- Schrand, C. M., & Walther, B. R. 2000. Strategic benchmarks in earnings announcements: the selective disclosure of prior - period earnings components. *The Accounting Review*, 75(2), 151-177.
- Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), 425-442.
- Sinha, N. R. 2016. Underreaction to news in the US stock market. *Quarterly Journal of Finance*, 6(2), 1650005.
- Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of accounting research*, 32(1), 38-60.
- Scott, W. R. 2015. *Financial accounting theory* (7. ed.). Toronto: Pearson.
- Tauringana, V., & Mangena, M. 2006. Complementary narrative commentaries of statutory accounts in annual reports of UK listed companies. *Journal of Applied Accounting Research*.
- Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3), 1139-1168.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. 2008. More than words: Quantifying language to measure firms' fundamentals. *The journal of finance*, 63(3), 1437-1467.
- Tsao, S., Lu, H. & Keung, E. C. 2018. Interim reporting frequency and the mispricing of accruals. *Accounting Horizons*, 32(3), 29-47.
- Tuomi, J. & Sarajärvi, A. 2018. Laadullinen tutkimus ja sisällönanalyysi (Uudistettu laitos.). Kustannusosakeyhtiö Tammi.

- Twedt, B., & Rees, L. 2012. Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy*, 31(1), 1-21.
- Xiao Wu, D., Yao, X. & Luan Guo, J. 2021. Is textual tone informative or inflated for firm's future value? evidence from chinese listed firms. *Economic Modelling*, 94, 513-525.
- Yee, K. K. 2004. Interim reporting frequency and financial analysts' expenditures. *Journal of Business Finance & Accounting*, 31(1-2), 167-198.
- You, H. & Zhang, X. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies*, 14(4), 559-586.
- Yuthas, K., Rogers, R., & Dillard, J. F. 2002. Communicative action and corporate annual reports. *Journal of Business Ethics*, 41(1), 141-157.

LIITE 3 ESIMERKKEJÄ SÄVYSANOISTA

Positiivisuus	Negatiivisuus	Epävarmuus
ACCOMPLISH	MISTAKE	ALTERATION
ACHIEVE	UNCOMPLETED	AMBIGUOUS
BETTER	BROKEN	ANTICIPATE
COLLABORATE	UNETHICALLY	APPROXIMATE
DILIGENT	RISKIER	CAUTIOUS
EASY	INADEQUACY	CONFUSING
ENHANCE	CONCERNS	DEVIATE
EXCEPTIONAL	MISMANAGING	DIFFER
FANTASTIC	ADVERSARY	DOUBT
GAIN	UNFAVORABLE	FLUCTUATE
GREAT	SACRIFICED	IMPRECISION
IMPRESS	WORRIES	INDEFINITE
IMPROVE	RECKLESS	POSSIBILITY
INNOVATE	UNPREDICTABLE	PRECAUTION
LEADING	DURESS	PRESUMPTION
OPTIMISTIC	WRITEDOWN	PROBABILITY
OUTPERFORM	UNFRIENDLY	RANDOM
PLEASED	DISGRACEFUL	REASSESSMENT
PROACTIVE	UPSET	RECONSIDER
REWARD	INACTION	RISK
SATISFACTION	ILLEGAL	RUMORS
STABILITY	DEMOLISH	SPECULATE
STRONG	DEVASTATION	SUGGEST
SUCCESS	STAGNATING	TENTATIVE
UPTURN	MALFUNCTIONS	UNPROVEN
VALUABLE	DIFFICULTY	VAGUE
WIN	DISADVANTAGED	VOLATILE

LIITE 4 KOODIESIMERKKEJÄ

Algoritmi tekstiosuuksien erottamiseen raporteista:

1. Etsitään jokin taulukko-osuutta kuvaava termi omalla rivillään ja poistetaan kaikki sen jälkeen oleva materiaali
2. Jaetaan jäljellä oleva osuus virkkeittäin (kohtuullisen hyvä tunnistaja `nltk.tokenize.sent_tokenizer`)
3. Jos virkkeessä on jokin yleinen sana (stopword), kuten `the`, `I`, `and` tai `a`, jatketaan tarkastelua.
4. Jos virke päättyy numeroihin ilman pistettä, mikä viittaisi taulukon riviin, se poistetaan, ellei seuraava (sivu)lause ala pienellä kirjaimella, mikä viittaisi virkkeen jatkumiseen.

5. Jos yli $\frac{3}{4}$ virkkeen sanoista alkaa isolla kirjaimella, mikä vaikuttaisi otsikolta, se poistetaan.
6. Jos kohdista 3-5 on päästy läpi, virke tallennetaan ja siirrytään kohtaan 3 seuraavan virkkeen osalta.

Epänormaalien tuottojen laskeminen:

Toistetaan jokaiselle aineiston raportille

1. Luetaan sisään csv-tiedostot, joissa tiedot yrityksen osakkeen päivittäisistä tuotoista sekä markkinoiden kehityksestä, myös SMB ja HML.
2. Haetaan tapahtumapäivän avulla estimointi-ikkunan tiedot ja tallennetaan ne 4x250 matriisiin (esimerkeissä df eli dataframe, datakehikko), jossa sarakkeet ovat yritys, markkina, SMB ja HML, riveinä 250 pörssipäivää.
3. Tehdään CAP- ja kolmen faktorin mallin sovitus ja palautetaan beta-kertoimet.

CAPM-beta suoralla laskulla: `df.corr()[hel_gi][yritys] * yritys_vol / markkina_vol`

Eli korrelaatiomatriisista otetaan markkinan (hel_gi) ja yrityksen korrelaatio, joka kerrotaan aikaisemmin lasketuilla tuottojen keskihajontojen suhteella.

Kolmen faktorin mallin beta-kertoimet sovittamalla regressiosuora: `linear_model.LinearRegression().fit(df[hel_gi,SMB,HML], yritys)`

Eli selittävät muuttujat hel_gi, SMB ja HML -tuotot, vaste yrityksen tuotot.

4. Verrataan mallien selitysasteita estimointi-ikkunan ajalta olevien toteutuneiden tuottojen osalta, korkeamman selitysasteen malli valitaan tapahtumaikkunaan.
5. Lasketaan toteutuneiden ja odotettujen tuottojen ero tapahtumaikkunassa luvun 4.2.1 kaavojen mukaisesti ja tallennetaan päivittäinen ero jokaiselta tapahtumapäivältä, joita voidaan käyttää kumulatiivisten epänormaalien tuottojen laskemiseen eri aikaikkunoilta.