



**KLUSTEROINTIMENETELMÄ ARVOSIJOITUSSTRATEGIANA LONTOON
PÖRSSISSÄ VUOSIEN 2010-2020 VÄLILLÄ**

Lappeenrannan–Lahden teknillinen yliopisto LUT

Kauppätieteiden kandidaatintutkielma

2022

Juho Lempinen

Tarkastaja: Tutkijatohtori Jyrki Savolainen

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppätieteet

Juho Lempinen

Klusterointimenetelmä arvosijoitusstrategiana Lontoon pörssissä vuosien 2010–2020 välillä

Kauppätieteiden kandidaatintyö

2022

41 sivua, 6 kuvaa ja 2 taulukkoa

Tarkastaja: Tutkijatohtori Jyrki Savolainen

Avainsanat: Arvosijoittaminen, Arvoanomalia, Klusterointi, Arvosijoitus strategia, K-means.

Tutkielmassa selvitetään, onko Lontoon osakemarkkinoilta pystynyt löytämään arvoanomaliaa käyttämällä K-means klusterointimenetelmää vuosien 2010–2020 välillä osakkeiden valinnassa. Tutkielmassa vertaillaan eri klusterimäärillä luotujen arvoportfolioiden raakatuottoa, sekä niiden riskikorjattua suoriutumista. Arvoportfoliot luodaan suodattamalla klusterointimenetelmällä tutkielman aineistosta ne osakkeet, joiden P/E, P/B ja P/CF tunnusluvut ovat matalimmat. Arvoportfolioiden tuottojen vertailun yhteydessä selvitetään myös, kuinka paljon valittu klusterimäärä vaikuttaa portfolion kokonaistuottoon.

Tutkielma koostuu kahdesta kokonaisuudesta. Ensimmäinen kokonaisuus tutkielmassa on kirjallisuuskatsaus. Toisena kokonaisuutena tutkielmassa on empiriaosia, joka suoritettiin kvantitatiivisena tutkimuksena.

Tutkimustulokset viittaavat, että käyttämällä klusterointimenetelmää Lontoon osakemarkkinoilla on pystynyt löytämään arvoanomalian tarkastellulla aikavälillä. Vertailtaessa kuudella eri klusterimäärällä luotuja arvoportfolioita, vain yksi luoduista portfolioista hävisi tuotolla mitattuna FTSE 100 Total Return- indeksille.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

School of Business and Management

Business Administration

Juho Lempinen

Clustering method for valuation strategy on the London Stock Exchange between 2010-2020

Bachelor's thesis

2022

41 pages, 6 figures and 2 tables

Examiner: Post-doctoral researcher Jyrki Savolainen

Keywords: Value investing, Value anomaly, Cluster, Value investing strategy, K-means.

This thesis examines whether it has been possible to find a value anomaly in the London stock market using the K-means clustering method between 2010 and 2020 for stock selection. This thesis compares the raw returns of value portfolios created with different numbers of clusters, as well as their risk-adjusted performance. Value portfolios are created by filtering the data from the thesis using the clustering method, the stocks with the lowest P/E, P/B, and P/CF ratios. In connection with the comparison of the returns of value portfolios, it is also determined how much the selected number of clusters affects the total return of the portfolio.

The thesis consists of two entities. The first entity in the thesis is a literature review. As a second entity, the dissertation has empirical parts, which were carried out as a quantitative study.

The research results suggest that by using the clustering method in the London stock market, it has been possible to find a value anomaly in the considered time period. When comparing value portfolios created with six different cluster numbers, only one of the created portfolios lost to the FTSE 100 Total Return index in terms of return.

LYHENTEET

FTSE	The Financial Times Stock Exchange
P/E	Price-to-Earnings; Yrityksen markkina-arvon suhde sen tulokseen
P/B	Price-to-Book; Yrityksen markkina-arvon suhde yrityksen omaan pääomaan
P/CF	Price-to-Cash Flow; Yrityksen markkina-arvon suhde sen liike-toiminnan kassavirtaan
NYSE	New York Stock Exchange
SOM	Self-Organizing Map; Itseorganisoituva kartta (Neuroverkko-malli)
CAP-malli	Capital asset pricing model

Sisällysluettelo

Tiivistelmä

Abstract

Lyhenneluettelo

1	Johdanto.....	1
1.1	Tutkielman tavoitteet ja tutkimuskysymykset	3
1.2	Tutkielman aineisto ja rakenne	4
2	Teoreettinen viitekehys	5
2.1	Tehokkaat markkinat.....	5
2.2	Arvoanomaliat.....	7
2.3	Moderni portfolioteoria.....	9
2.3.1	CAP- malli	11
2.4	K-means klusterointi	14
2.5	Jensenin alfa	16
2.6	Sharpen luku.....	17
3	Aineisto ja tutkimusmenetelmät	18
3.1	Tutkielman aineisto	18
3.2	Tutkimusmenetelmät ja portfolioiden muodostaminen	19
4	Tutkimustulokset	21
4.1	Klusterointimenetelmällä luotujen portfolioiden suoriutuminen	21
4.2	Arvoportfolioiden riskikorjattu suoriutuminen	25
4.3	Tulosten pohdinta.....	27
5	Johtopäätökset ja yhteenveto	29
5.1	Tutkimustulosten luotettavuuden arviointi ja jatkotutkimukset.....	31
	Lähteet	33

1 Johdanto

Arvosijoittaminen on rahoitusmaailmassa suosittu tutkimusten aihe. Useat tutkijat (Basu, 1977; Stattman, 1980; Dessanaike & Lim, 2010; Fama & French, 1998; Capaul, Rowley & Sharpe, 1993; Ball, Gerakos, Linnainmaa & Nikolaev, 2020) ovat havainneet, että arvo-osakkeet ovat suoriutuneet hyvin markkinaan nähden, joka on tukenut arvoanomalian syntymistä. Arvoanomalia liittyy keskeisesti Faman (1970) esittämään tehokkaiden markkinoitten teoriaan. Anomaliassa rationaaliset sijoittajat pyrkivät saamaan ylituottoja markkinoilta etsimällä sellaisia osakkeita, jotka ovat markkinan epätehokkuuden johtuen aliarvostettuja. Olettaen samalla, että tehokkaat markkinat korjaavat aliarvostettujen osakkeiden arvostuksen oikealle tasolle. (Fama & French, 1970; Fama & French, 1998) Tässä tutkielmassa keskitytään arvoanomaliaan arvosijoittamisen kautta. Tutkielmassa arvo-osakkeet valitaan K-means algoritmia hyödyntäen. Algoritmi muodostaa aineistosta klustereita, joista muodostetaan tutkielmassa analysoitavat portfoliot.

Arvosijoittamisen strategian suurelle yleisölle esitteli vuonna 1934 Graham ja Dodd julkaisemassaan kirjassa *Security Analysis*. Teoksessaan he pitivät sijoituksen menestymisen mittarina yhtiön laatua. He painottivat tunnuslukujen ja fundamenttien, kuten positiivisen liiketoiminnan kassavirran, oman pääoman kirjanpidollisen arvon ja yhtiön tuloksen tekokyvyn tärkeyttä laadukkaiden yhtiöiden löytämisessä. (Graham & Dodd, 1934) Grahamin ja Doddin (1934) mukaan näiden periaatteiden valossa arvo-osake voidaan tunnistaa tunnuslukujen, kuten P/B:n, P/E:n sekä P/CF:n avulla. P/B- tunnusluku (engl. Price-to-book ratio) kuvaa yrityksen osakkeen hinnan suhdetta sen osakekohtaiseen omaan pääomaan. Tunnusluku kertoo, onko yrityksen osakekohtainen kirjanpidollinen arvo suurempi tai pienempi kuin yrityksen osakekohtainen hinta. (Bodie, Kane, Marcus, 2005) P/E tunnusluvulla taas (engl. Price-to-earnings ratio) tarkoitetaan yrityksen osakkeen hinnan ja sen luoman osakekohtaisen tuloksen välistä suhdetta. Tunnusluku kertoo, kuinka paljon sijoittaja joutuu maksamaan yhdestä yrityksen tuottamasta eurosta ja kuinka kauan yrityksellä kestää tuottaa voittoa siihen sijoitetun summan verran. (Bodie *et al.*, 2005) Vastaavasti P/CF tunnusluvulla kuvataan

yrittäjän osakkeen hinnan ja sen liiketoiminnasta syntyvän kassavirran välistä suhdetta. Tunnusluku kertoo, kuinka kauan yrityksellä kestää tuottaa sen oman markkina-arvonsa verran kassavirtaa, jos sen markkina-arvo pysyy muuttumattomana. (Bodie *et al.*, 2005)

Akateeminen arvo-osakkeen määrittäminen tunnusluvuilla poikkeaa kuitenkin hieman Grahamin ja Doddin määritelmästä. Akateemisessa maailmassa arvo-osakkeet määritellään perinteisesti P/E- ja P/B tunnuslukujen mukaan, joita suhteutetaan osakkeen markkina-arvoon. Muodostuneen suhdeluvun avulla pyritään arvioimaan, onko osake ali- vai yliarvostettu markkinoilla. (Nicholson, 1960; Stattman, 1980) Perinteinen arvosijoitusstrategia perustuu osakkeiden ali- ja yliarvostuksen tulkitsemiseen ja määrittämiseen. Arvosijoittamisessa on tyypillistä, että sijoittaja pyrkii valitsemaan portfolioonsa yhtiöitä, jotka ovat tunnuslukujen valossa aliarvostettuja. Eli sellaisia osakkeita, joiden tunnusluvut ovat matalia. (Hanso, 2013; Graham & Dodd; 1934)

Tämä tutkielma tarjoaa modernin ja vaihtoehtoisen tavan harjoittaa arvosijoittamista tunnuslukujen perusteella. Osakemarkkinat ympäri maailmaa ovat hyvin suuret, jonka seurauksena eri osakkeiden tunnuslukujen seulominen on aikaa vievää. K-means algoritmin avulla sijoittaja pystyy luomaan suuresta osakemassasta pienempiä kokonaisuuksia. Aiemman kirjallisuuden mukaan K-means klusterointi on tehokas menetelmä osakkeiden valintaan ja portfolioiden muodostamiseen. (Cheong, Kim, Byun, Oh ja Kim, 2017; Nandan, Mahantyn ja Tiwarin, 2010) Cheong *et al.*, (2017) pystyivät saavuttamaan tutkimuksessaan 24,12 % tuoton muodostamalla portfolioit K-means klusterointimenetelmän avulla. Heidän saavuttama tuotto on huomattavasti osakemarkkinan keskiarvo tuottoa suurempi. Tämän vuoksi on mielenkiintoista nähdä, millaisia tuottoja ja riskikorjattuja tuottoja sijoittaja olisi voinut saada Lontoon osakemarkkinoilla vuosien 2010–2020 välisellä ajalla käyttämällä K-means menetelmää.

Tutkielman aihe on tärkeä juuri nyt, koska anomaliatutkimuksissa käytetty osakedata on hyvin vanhaa. Osakemarkkinat ovat muuttuneet huomattavasti 2000-luvun jälkeen, erityisesti teknologian kehityksen myötä. Teknologian kehityksen myötä myös kasvu osakkeisiin sijoittaminen on yleistynyt huomattavasti, ja siksi usein kuulee, että arvosijoittaminen sekä

tunnuslukujen tulkita on vanhanaikaista. Tämä herättää kysymyksen arvo-osakkeiden suoriutumisesta ja arvoanomaliasta, joihin tässä tutkielmassa perehdytään.

Tämä tutkielma tehdään tutkimustyönä Valuatum Oy yritykselle. Työn tarkoituksena on kerätä tietoa K-means algoritmin suoriutumisesta, jota yritys pystyy hyödyntämään uuden tuotteen kehittämisessä rahastoille, varallisuudenhallintaan ja yksityissijoittajille. Tutkielman myötä yritys pystyy saamaan laajan kokonaiskuva siitä, miten K-means klusterointimenetelmä suoriutuu osakkeiden valinnassa ja portfolioiden luomisen tukena. Tutkielman johtopäätökset pääluku tarjoaa yritykselle ideoita, kuinka tutkielmassa käytettävää menetelmää voisi kehittää eteenpäin

1.1 Tutkielman tavoitteet ja tutkimuskysymykset

Tämän tutkielman tavoitteena on selvittää, onko arvosijoittamisella pystynyt saamaan riskikorjattuja ylituottoja suhteessa markkinaan muodostamalla arvo-osake portfolioit käyttämällä automatisoitua K-means algoritmia. Portfolioit muodostetaan P/E, P/B ja P/CF tunnuslukuja käyttäen, joiden soveltuvuus arvo-osakkeiden valintaan saa tukea aikaisemmasta kirjallisuudesta. (Basu,1997, Fama ja French, 1998; Dessanaike & Lim, 2010) Tutkielman tavoitteen ja aihetta käsittelevän kirjallisuuden perusteella tutkielman pääkysymykseksi valikoitui:

- Voidaanko FTSE 100:sta löytyä arvoanomaliaa vuosien 2010–2020 välillä käyttämällä K-means menetelmää?

Lisäksi tutkielmassa on tarkoitus testata eri klustereiden määrää ja tarkastella niiden muodostamia tuottoja. Sen tarkoitus on tuottaa informaatiota siitä, kuinka monta klusteria sijoittajan kannattaa valita suoriutuakseen parhaiten. Tästä syystä tutkielmassa pyritään vastaamaan seuraavaan kysymykseen:

-Kuinka paljon valittu klustereiden kokonaismäärä olisi vaikuttanut kokonaistuottoihin vuosien 2010–2020 aikana?

1.2 Tutkielman aineisto ja rakenne

Tutkielman aineistona käytetään Lontoon pörssissä olevan FTSE-100 indeksissä vuosien 2010–2020 välisenä aikana olevia osakkeita. FTSE-100 indeksi sisältää 100 suurinta Lontoon pörssissä listattua yhtiötä. Aineiston valinta toimii myös tutkielman maantieteellisenä rajauksena. Tutkielmassa vertailukohtana käytetään FTSE-100 Total Return indeksii, joka sisältää osinkojen tuotot. Tämä mahdollistaa sen, että kokonaistuotot ovat vertailukelpoisia keskenään, eivätkä ne vaadi yhdenvertaistuksia. Tutkielmassa portfoliot muodostetaan klustroimalla osakkeet P/E:n, P/B:n ja P/CF:n mukaan K-means algoritmilla. Osakkeiden klusterointi tehdään uudestaan joka vuosi tarkasteluperiodin aikana, jolloin indeksistä poistettut ja kriteereihin sopimattomat yhtiöt poistetaan aineistosta. Portfolioihin ei valita tutkielmassa sellaisia yhtiöitä, joiden tulos tai kassavirta on negatiivinen, jolloin saadaan laadukkaita mahdolliset arvoportfoliot. Tutkielmassa tuottojen laskennassa ei oteta huomioon, mitään osakkeiden myynnistä tai ostamisesta, tai muista toimenpiteistä aiheutuvia kuluja.

Tutkielman rakenne koostuu viidestä pääluvusta. Ensimmäisessä pääluvussa eli johdannossa oli tarkoitus johdattaa lukija tutkielmaan. Johdannossa esitellään lukijalle tutkimuksen aihe, sen tavoitteet ja aikaisempia aiheeseen liittyviä tutkimuksia. Toisessa pääluvussa esitellään tutkielman kannalta relevantti teoreettinen viitekehys. Viitekehys pitää sisällään tehokkaiden markkinoiden hypoteesin, modernin portfolioteorian ja siihen liittyvän CAP- mallin. Toinen pääluku pitää sisällään myös tutkielmassa käytetyn K-means klusteroinnin esittelyn, ja tutkielman kannalta keskeiset suoriutumismittarit ja tunnusluvut. Kolmannessa pääluvussa kuvaillaan lukijalle tutkielmassa käytettävää aineistoa, tutkimusmenetelmiä ja arvoosake portfolioiden muodostamista. Tutkielman neljännessä pääluvussa esitellään tutkimuksen tulokset ja analysoida niitä. Viidennessä pääluvussa esitellään tutkimustuloksista tehdyt johtopäätökset ja samalla arvioidaan tutkimustulosten luotettavuutta.

2 Teoreettinen viitekehys

Tutkielman teoreettinen viitekehys sisältää kuusi alalukua, jotka ovat keskeisiä tutkielman kannalta. Viitekehysten ensimmäisessä alaluvussa käsitellään tehokkaiden markkinoiden hypoteesia, joka on keskeinen teoria arvoanomalian kannalta. Toisessa alaluvussa käsitellään aikaisempia tutkimuksia, jotka puoltavat arvoanomalian olemassaoloa. Aiemmat tutkimukset toimivat myös tukena sille, miksi tutkielmaan on valittu juuri kyseisessä kappaleessa esiintyvät taloudelliset tunnusluvut. Neljännessä kappaleessa esitetään moderni portfolio teoria ja CAP- malli. Tutkielmassa oletuksena on, että portfoliot muodostetaan K-means algoritmilla, eikä näin ollen tässä tutkielmassa kiinnitetä huomiota portfolion arvopapereiden keskinäiseen korrelaatioon. Viidennessä alakappaleessa esitetään K-means algoritmin peruseriaate ja toiminta. Lisäksi esitetään aikaisempia tutkimuksia K-means algoritmin toiminnasta. Viimeisessä kappaleessa esitellään riskikorjatun tuoton mittarit, jotka ovat tässä tutkimuksessa Sharpen luku ja Jensenin alfa.

2.1 Tehokkaat markkinat

Tunnettu tehokkaiden markkinoiden teoria perustuu Faman 1970 luvulla esittämään hypoteesiin. Hypoteesin mukaan osakkeiden nykyiset hinnat heijastavat kaiken markkinoilla olevan relevantin informaation, osakkeisiin kohdistuvat tulevaisuuden tuotto-odotukset ja valitsevat riskit. (Fama, 1970) Hypoteesi perustuu oletukseen, jonka mukaan sijoittajat eivät voi saada ylituottoja markkinoilta, koska kaikilla toimijoilla on käytössään sama informaatio. Tästä syystä aikaisemmalla informaatiolla ei ole mahdollista saada informaatio hyötyjä markkinoilta, koska sijoittajat toimivat teorian mukaan rationaalisesti. Sijoittajien rationaalisen käytöksen ansiosta markkinoiden hinnanmuodostus on tehokasta, koska pääomat kohdentuvat tehokkaasti ja informaatio hinnoitellaan aina osakkeen hintaan oikein. (Fama, 1970; Fama 1995) Tehokkaat markkinat voidaan Faman (1970) mukaan jakaa kolmeen eri tasoon sen mukaan, kuinka paljon informaatiota on saatavilla, ja kuinka hyvin se heijastuu osakkeiden nykyisiin hintoihin.

Tehokkaiden markkinoiden ensimmäistä tasoa kutsutaan heikosti tehokkaiksi markkinoiksi. Heikosti tehokkaiden markkinoiden määritelmän mukaan nykyinen osakkeen hinta pitää sisällään kaiken historiallisen informaation. Tällöin sijoittajan ei ole mahdollista ennustaa osakkeen tulevaisuuden kehitystä historiallisen tiedon tai uutisten avulla. Tehokkaiden markkinoiden toista tasoa kutsutaan puolivahvasti tehokkaiksi markkinoiksi. Puolivahvasti tehokkailla markkinoilla osakkeiden nykyinen hinta pitää sisällään heikosti tehokkaan markkinan ominaisuudet. Puolivahvasti tehokkaat markkinat pitävät heikosti tehokkaan markkinan ominaisuuksien lisäksi myös kaiken julkisesti saatavilla olevan informaation, kuten tulosjulkistukset ja muut yhtiötä koskevat uutiset. Markkinoiden ollessa puolivahvasti tehokkaat sijoittajat eivät pysty saamaan ylituottoa markkinoilta löytämällä väärin hinnoiteltuja osakkeita analysoimalla yhtiöiden taloustietoja. Viimeistä tasoa kutsutaan vahvasti tehokkaiksi markkinoiksi. Vahvasti tehokkailla markkinoilla osakkeiden nykyinen hinta pitää sisällään kahden aiemman tason ominaisuudet, sekä niiden lisäksi kaiken sisäpiirin tiedon. Tällaisilla markkinoilla kukaan sijoittaja tai muut toimija ei pysty saamaan markkinaa ylittävää tuottoa koska sisäpiirissä syntyvä päätös heijastuu välittömästi osakkeen hintaan. (Fama, 1970)

Tehokkaiden markkinoiden yhteydessä tulee ottaa huomioon myös osakkeen tuoton satunnaiskävelyn teoria. Teorian mukaan osakkeen tulevaisuuden tuotot ovat satunnaisia, eikä niitä pysty ennustamaan koska tehokkaiden markkinoiden mukaan kaikki relevantti historiallinen tieto on huomioitu osakkeen hinnassa. Tästä syystä osakkeiden hinnat muuttuvat vain silloin, kun uutta informaatiota ilmestyy markkinoille. Uuden markkinoille tulevan informaation tulee olla sellaista, jota sijoittajat eivät ole pystyneet ennustamaan. Jos uusi informaatio ei täytä ennustamattomuuden ehtoa, se on jo heijastunut osakkeen nykyiseen hintaan. (Knüpfer & Puttonen, 2018)

Sijoittajien reagointi uuteen informaation voi olla välitöntä tai hidasta riippuen informaation laajuudesta. Sijoittajien välittömässä reagoinnissa uusi informaatio hinnoitellaan välittömästi osakkeen nykyiseen hintaan, jolloin osakkeen hinnan tulisi vastata täydellisesti saadun informaation vaikutusta. Tällöin kukaan sijoittaja ei voi hyödyntää viivästyneitä

hintareaktioita sijoituspäätöksissään. Jos hinnat reagoisivat täydellisesti täydessä laajuudessaan saatuun informaatio, ei teorian mukaan markkinoilla tulisi silloin olla väärin hinnoiteltuja osakkeita informaation suhteen. (Kallunki & Niemelä, 2012, 244)

Teoriassa sijoittajia on pidetty rationaalisina toimijoina markkinoilla, ja he pyrkivät aina maksimoimaan heidän tuottoensa tekemättä virheitä. (Fama, 1995) Malkiel (2003) argumentoi tutkimuksessaan tätä väitettä vastaan. Hänen mukaansa markkinoilla olevat sijoittajat eivät tee rationaalisia päätöksiä, vaan enemmänkin epärationaalisia päätöksiä. Nämä epärationaaliset päätökset johtavat markkinoiden epätehokkuuksiin ja hinnoitteluvirheisiin. Tehokkaiden markkinoiden teoria on saanut vuosien saatossa paljon kritiikkiä. Fama (1991) argumentoi uudemmassa tutkimuksessaan noin kaksikymmentä vuotta aiemmin esittämänsä tehokkaiden markkinoiden hypoteesia vastaan. Hänen mielestään markkinat eivät voi olla täysin tehokkaat koska markkinoilta on löytynyt useita tunnuslukuja, joiden mukaan markkinoilla on voinut saada ylituottoja. Sijoittajat voivat hänen mielestään myös tulkita taloudellisia tai poliittisia tapahtumia väärin rationaalisuuden puutteen vuoksi, joka aiheuttaa markkinoille epätehokkuuksia. (Fama, 1991) Tutkielman seuraava alakappale esittelee aikaisempia tutkimuksia, joissa ylituottoa on saatu eri tunnuslukujen avulla.

2.2 Arvoanomaliat

Arvoanomalialla tarkoitetaan tilannetta, jossa markkinoille syntyy epäsäännöllisyyksiä tai poikkeamia, eikä näitä tapahtumia pystytä selittämään perinteisillä malleilla. Arvoanomalian tilanteessa markkinoilla olevat arvo-osakkeet ovat toimineet aiemmin mainitun tehokkaiden markkinoiden teorian vastaisesti. Tämän kaltaisessa tilanteessa arvo-osakkeet ovat suoriutuneet markkinoita paremmin eli tuottaneet ylituottoja, tai ne ovat toimineet CAP-mallin odotusten vastaisesti. (Frankfurter & McGoun, 2001; Elze, 2012, 633)

Arvo-osakkeiden suoriutumista eri tunnusluvuilla on tutkittu rahoituksen maailmassa huomattavan paljon. Basu (1977) on yksi ensimmäisistä tutkijoista, jotka testasivat P/E luvun suoriutumista osakemarkkinoilla. Tutkimuksessaan hän testasi P/E-anomalian

olemassaoloa. Hän tutki tutkimuksessaan yli 1400 NYSE:ssä listatun osakkeen P/E lukua ja sen suoriutumista vuosien 1957–1971 välillä. Tutkimuksessaan hän muodosti viisi portfolioa P/E luvun tason mukaan, joiden suoriutumista hän vertasi markkinaportfolioon. Hänen tutkimustulostensa mukaan matalan P/E:n portfolioit suoriutuivat absoluuttisesti paremmin ja tuottivat korkeamman riskikorjatun tuoton kuin korkean P/E:n osakkeet.

P/E-anomaliaa tukee myös Anderssonin ja Brooks (2006) tutkimus. He tutkivat P/E tunnusluvun suoriutumista Ison-Britannian osakemarkkinoilla vuosien 1975–2003 välillä. Tutkimuksessaan he argumentoivat, että liian lyhyeltä aikaväliltä lasketut tuotot antavat aliarvioidun kuvan yrityksen tuloksenteo kyvystä. Tästä syystä he muodostivat osakeportfolioit P/E:n tason mukaan, jonka he olivat laskeneet jakamalla yhdestä kahdeksaan vuoden periodin tuottojen summalla sen osakekurssin. Heidän tutkimustulostensa mukaan matalan P/E:n portfolioit tuottivat paremmin kuin korkean P/E:n portfolioit riippumatta siitä, kuinka pitkältä ajalta tuotot olivat laskettu. Matalan P/E:n portfolioin keskimääräinen tuotto tutkitavalta ajanjaksolta oli 26,68 %, kun vastaavasti korkean P/E:n portfolioiden 13,99 %. (Andersson & Brooks, 2006)

Kirjallisuuden mukaan arvoanomaliaa on löytynyt myös P/B tunnusluvun avulla (Stattman, 1980). Fama ja French (1998) vertailivat tutkimuksessaan arvo-osakkeiden suoriutumista kasvu osakkeisiin, sekä markkinaan. Heidän tutkimuksensa mukaan arvo-osakkeet olivat tuottaneet paremmin kuin kasvu osakkeet ja markkina vuosien 1975–1995 välillä. Heidän tutkimustuloksensa tukee arvoanomalian olemassaoloa. Suurin huomio kuitenkin heidän tutkimuksissaan oli se, että matalalla P/B tunnusluvulla valituilla arvo-osakkeilla olisi voinut saada ylituottoja kahdessatoista kolmestatoista tutkitusta markkinasta tarkasteluperiodiin aikana. Capaul, Rowley ja Sharpe saivat vuonna 1993 tekemässään tutkimuksessaan myös samankaltaisia tuloksia kuin Fama ja French. Capaul, Rowley ja Sharpe (1993) osoittivat tutkimuksessaan, että Ison-Britannian osakemarkkinoilla oli pystynyt saamaan ylituottoja muodostamalla portfolioit matalalla P/B luvulla.

Markkinoilta on löytynyt tutkimuksien mukaan arvoanomaliaita myös P/CF tunnusluvulla. Tunnusluku poikkeaa perinteisistä akateemisesti määritellyistä arvo-osakkeiden

tunnuslukuista. Dessanaike ja Lim (2010) tutkivat Ohlson- ja RIM valuaatiomalli strategian suoriutumista iso-britannian osakemarkkinoilla vuosien 1987–2001 aikana. Mallit pyrkivät ennustamaan yrityksen tulevaisuuden tuottoja ja käyttämään hyväksi paluu keskiarvoon ilmiötä. Tutkimuksessaan he vertasivat RIM- ja Ohlson- mallien antamia tuloksia perinteiseen tunnusluku strategiaan, johon he ottivat mukaan P/B, P/E ja P/CF tunnusluvut. Heidän tutkimustuloksensa osoittavat, että hienostuneet mallit suoriutuivat vain hieman P/CF tunnusluku strategiaa paremmin. Kuitenkin molempien strategioiden mukaan tehdyt arvoportfoliot suoriutuivat kasvuportfolioita paremmin. (Dessanaike & Lim, 2010)

P/CF tunnusluku saa tukea myös Faman ja Frenchin (1998) tutkimuksesta. Samassa tutkimuksessa, jossa he löysivät P/E ja P/B anomalian, he löysivät myös P/CF anomalian. Heidän tutkimustulostensa mukaan matalalla P/CF tunnusluvulla luodut arvoportfoliot olivat suoriutuneet paremmin kuin korkean P/CF:n kasvuportfoliot vuosien 1975–1995 välillä. Arvoportfolioiden tuotto tarkasteluajanjaksolla oli keskimäärin 15,9 %, kun kasvuportfolioiden tuotto oli vain 9,27 %. Lisäksi tutkimuksessa selviää, että matalan P/CF:n tunnusluvun portfoliot olivat tuottaneet iso-britannian markkinoilla keskimäärin 18,41 % tarkasteluperiodilla. Kun vastaavasti korkean P/CF tunnusluvun kasvuportfoliot olivat tuottaneet keskimäärin 14,51 % ja markkina 15,33 %. (Fama & French, 1998)

Kappaleessa esitettyjen aikaisempien tutkimusten mukaan P/E, P/B ja P/CF tunnusluvuilla oli saatu myönteisiä tuloksia arvoanomalian olemassaolosta. Tästä syystä ne soveltuvat arvo-osakkeiden etsintään ja arvoportfolioiden muodostamiseen tutkielman aineistosta. Seuraavassa alakappaleessa perehdytään siihen, miten portfoliot tulisi muodostaa portfolioteorian perusteella.

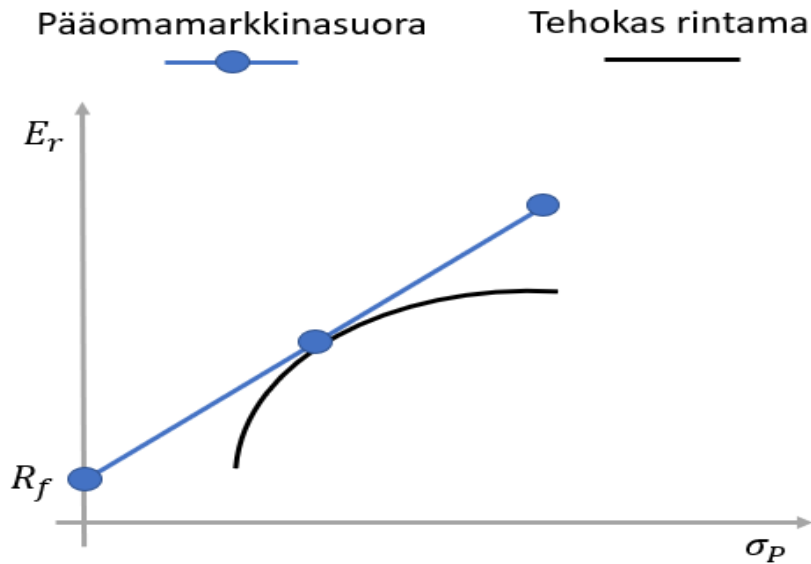
2.3 Moderni portfolioteoria

Moderni portfolioteoria on nykyäänkin yksi sijoitusmaailman tunnetuimmista portfolion hallintaa liittyvistä teorioista, jonka Markowitz esitteli suurelle yleisölle vuonna 1952. Hänen luoman teoriasensa yksi keskeinen idea on portfolion hajauttamisesta eri

omaisuusluokkien kesken saatava hyöty. Teorian mukaan portfolioon tulisi valita sellaisia arvopapereita, joiden keskeinen korrelaatio on mahdollisimman lähellä nollaa. Se mahdollistaa portfolioissa olevien arvopapereiden vastakkaisen liikehdinnän, jolloin yksittäiset arvopaperit suojaavat toinen toistaan. Tämänkaltaisen arvopapereiden hajauttamien mahdollistaa portfolioon kohdistuvan kokonaisriskin alenemisen vähentämättä portfolioissa olevien arvopapereiden tuotto-odotusta. (Markowitz, 1952)

Markowitzin (1952) teorian mukaan sijoittajan on tärkeää kiinnittää huomiota portfolioon sisältöön ja luoda siitä omaan riskinotto-kykyyn soveltuva. Markowitz (1952) piti kokonaisuuden kannalta tärkeänä sitä, kuinka suuret painoarvot yksittäiset osakkeet saavat portfolioissa. Yksittäisten osakkeiden painoarvoilla on hänen mukaansa huomattava vaikutus koko portfolioon tuottoon ja keskihajontaan. Painottamalla korkean riskin ja korkean tuoton osakkeita portfolioon kokonaisriski kasvaa, joka samalla kasvattaa sen kokonaistuottoa. Vastaavasti sijoittaja voi laskea portfolioon kokonaisriskiä painottamalla enemmän matalan tuoton ja matalan riskin osakkeita. (Markowitz, 1952)

Markowitz loi esittämässään teoriassaan käsitteen tehokkaasta rintamasta, josta muodostui yksi teorian keskeisimmistä ideoista optimaalisen portfolioon valitsemiseen. Tehokkaalla rintamalla voidaan teoria mukaan kuvata optimaalista odotetun tuoton ja volatilititeetin suhdetta. (Markowitz, 1952) Kuvassa 1 esitetty tehokas rintama (Mukaiillen Bodie *et al.*, 2005) koostuu tuottoriski suhteeltaan optimimaalisista portfolioista, jotka sisältävät varianssin minimoivia riskillisiä arvopapereita. (Markowitz, 1952; Perold, 2004) Kun tarkasteluun otetaan mukaan myös riskittömät omaisuuserät, saadaan Kuvassa 1 oleva pääomamarkkinasuora. Sen tangentti pitää sisällään kaikki ne keskiarvo-variassitehokkaat portfolioit, jotka sisältävät riskittömän koron komponentteja. (Fama *et al.*, 2004) Lisäksi niiden riski ja tuotto ovat optimaalisella tasolla eli ne maksimoivat tuoton tietyllä riskitasolla. Tämä tarkoittaa sitä, että jokainen pääomamarkkinasuoralla oleva piste on tehokas. Jos portfolio sijaitsee tehokkaalla rintamalla sekä pääomamarkkinasuoralla, sisältävät nämä portfolioit riskillisiä ja riskittömiä arvopapereita. Jos pääomamarkkinasuoraan otetaan mukaan myös lainaaminen, voidaan tilannetta kutsua nimellä Tobinin eroteltavuusteoreema (Bodie *et al.*, 2005; Fama *et al.*, 2004) Kuvassa 1 E_r edustaa odotettua tuottoa, R_f riskitöntä tuottoa ja σ_p portfolioon volatilititeettiä.



Kuva 1. Tehokas rintama ja pääomamarkkinasuora (Mukaiillen Bodie *et al.*, 2005)

Kuvassa 1 aiemmin mainittu optimaalinen tuottoriski suhde voidaan löytää kohdasta, jossa tehokas rintama kohtaa pääomamarkkinasuoran. Optimaalista kohtaa voidaan kutsua myös tangentiportfolioksi. Portfoliot, jotka ovat optimaalisen pisteen yläpuolella ja jäävät käyrälle ovat tehokkaita ja niistä sijoittajan tulisi valita omaan tuotto- ja riskiprofiiliin sopiva portfolio. Sijoittajan ei kannata valita sellaisia portfolioita, jotka jäävät tehokkaan rintaman optimaalisen kohdan alapuolelle. Alapuolelle jäävän portfolion tuottoriski suhde on tehoton. Tällöin sijoittajan portfolion riski kasvaa, mutta tuotto-odotus laskee. (Knüpfer *et al.*, 2018)

2.3.1 CAP- malli

Nykyistä CAP- mallia voidaan luonnehtia monen tutkijan tutkimuksen lopputuloksena. Ensimmäisen version CAP-mallista toi suuren yleisön tietoisuuteen Sharpe (1964). Kuitenkin Lintnerillä (1965) ja Blackillä (1972) on ollut suuri vaikutus nykyisen CAP- mallin

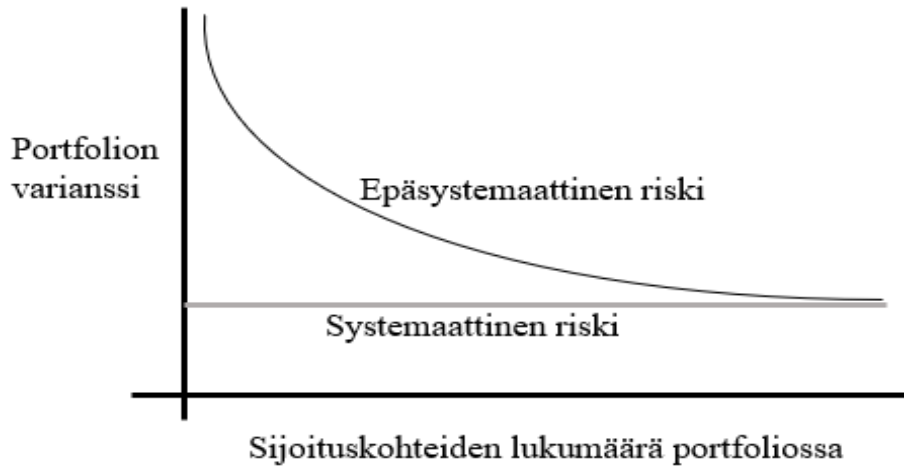
kehitykseen. Sharpen luomalla CAP-mallilla pyritään kuvaamaan riskin ja tuoton välistä suhdetta. (Knüpfer *et al.*,2018) Teorian mukaan CAP- mallissa oletuksena on, että markkinoilla on täydellinen kilpailu, jossa sijoittajat pyrkivät maksimoimaan oman tuottonsa toimien rationaalisesti optimaalisen riskinoton puitteissa. Teoria olettaa myös, että sijoittajat pystyvät myymään lyhyeksi sijoituskohdetta ja jakamaan sen osiksi, sekä likvidoimaan kohteen jokaisena ajankohtana ilman kustannuksia. Lisäksi kaikki sijoittajat pystyvät ottamaan velkaa rajoituksetta vallitsevalla riskittömällä korolla, ja heillä on käytössään samanaikaisesti tehokkaiden markkinoiden määritelmän mukainen uusi informaatio. Teorian odotukset pyrkivät luomaan yksinkertaistetun reaali maailman, jossa sijoittajat sijoittavat samalle aika periodille ja heillä on homogeeniset oletukset. (Bodie *et al.*, 2005)

CAP- mallia voidaan käyttää myös arvopapereiden tuotto-odotuksen laskemiseen. Sharpen (1964) mukaan CAP-malli voidaan kuvata seuraavalla kaavalla:

$$E(r_i) = r_f + \beta_i[E(r_m) - r_f] \quad (1)$$

Jossa $E(r_i)$ kuvaa arvopaperin i tuotto-odotusta ja r_f kuvaa riskitöntä korkoa. Kaavassa riskikomponenttia edustaa β_i , joka kuvaa sijoituskohteen i beta-kerrointa. Kaavassa $E(r_m)$ kuvaa markkinaportfolion tuotto-odotusta. Yhtälössä termi $E(r_m) - r_f$ kuvaa riskipreemiota.

Muodostetun portfolion kokonaisriski voidaan jakaa kahteen eri osaan. Systemaattiseen riskiin ja epäsystemaattiseen riskiin. Epäsystemaattista riskiä syntyy sellaisista tekijöistä, jotka ovat yrityskohtaisia, eikä niillä ole vaikutusta muiden portfoliossa olevien osakkeiden tuottoihin. Sijoittajat pystyvät vähentämään epäsystemaattista riskiä hajauttamalla portfolionsa modernin portfolioteorian mukaisesti useisiin eri osakkeisiin tai omaisuusluokkiin. Mitä pienempi uuden osakkeen korrelaatio on salkussa jo olevien osakkeiden kanssa, sitä suurempi vaikutus sillä on portfolion varianssiin. (Knüpfer *et al.*,2018) Portfolion hajauttamisen vaikutusta sen varianssiin esitetään kuvassa 2.



Kuva 2. Systemaattinen riski ja epäsystemaattinen riski (Mukaillen Knüpfer *et al.*,2018)

Vastaavasti systemaattisella riskillä eli markkinariskillä tarkoitetaan sellaista riskiä, joka on vallitsevasta taloudesta ja markkinoista riippuvainen. Se vaikuttaa kaikkiin markkinoilla oleviin osakkeisiin samanaikaisesti, mutta systemaattisen riskin tuoman markkinariskin vaikutus voi vaihdella toimialojen välillä. Systemaattista riskiä sijoittaja ei pysty hajauttamaan pois, joten se jää vallitsevaksi minimi riskitasoksi portfolioon. Tästä syystä systemaattinen riski on otettava huomioon tuottovaatimuksen määrittämisessä, kuten aiemmin esitetystä CAP- mallin yhtälössä on tehty β_i eli beta-kertoimella. (Knüpfer *et al.*,2018) Beta-kerroin edustaa yhden osakkeen systemaattista riskiä. β_i lasketaan seuraavalla kaavalla:

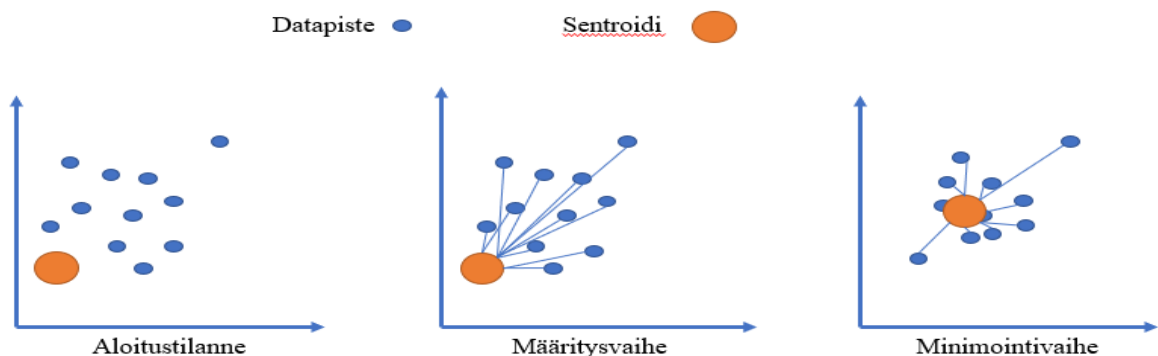
$$\beta_i = \frac{\text{cov}(R_i, R_m)}{\text{var}(R_m)} \quad (2)$$

Jossa kaavan osoittajana oleva $\text{cov}(R_i, R_m)$ kuvaa sijoituskohteen i tuoton ja markkinaportfolion R_m tuoton kovarianssia. Kaavan nimittäjä kuvaa markkinaportfolio R_m tuoton varianssia.

2.4 K-means klusterointi

K-means klusterointi menetelmän yhdistäminen osakkeiden valintaan on huomattavasti uudempi ilmiö kirjallisuudessa, kuin perinteinen osakkeiden valitsemisen tutkiminen. K-means algoritmin suuren yleisön tietoisuuteen toi Forgy vuonna 1965 ja Lloyd vuonna 1982. (Forgy, 1965; Lloyd 1982) K-means algoritmiä voidaan kutsua myös Lloyd-Forgy algoritmiksi. Se on menetelmänä yksi vanhimmista ja useimmiten käytetyistä ohjaamattoman oppimisen (engl. Unsupervised learning) klusterointialgoritmeista. (Gan & Ng, 2017) Klusterointi menetelmän tarkoitus on paljastaa luokittelurakenteita sekä tunnistaa homogeenisiä ryhmiä tai klustereita jostakin reaali maailmasta kerätystä datasta, ja yhdistää saman luokan havainnot yhdeksi klusteriksi. (Affonso, Magela, Dias & Pinto, 2021; Gan & Ng, 2017) Tässä tutkielmassa reaali maailman datana käytetään yhtiöiden taloudellisesta tiedosta muodostettuja tunnuslukuja.

K-means algoritmi toimii yksinkertaisella kierrolla. Se ottaa syötteenä käytettävien sentroidien lukumäärän. Sentroidilla tarkoitetaan klusteroinnissa sitä pistettä, joka sijaitsee datapiste joukon kaikkien pisteiden keskiarvolla. Eli, se määrittää kuinka monta klusteria algoritmin halutaan muodostavan. Kun klustereiden määrä on määritelty, algoritmi sijoittaa sentroidit satunnaiseen paikkaan datapisteiden vektoriavaruudessa. Algoritmin klusterin muodostaminen voidaan jakaa kahteen eri vaiheeseen. Määrittämisvaiheeseen ja minimointivaiheeseen, joita esitetään kuvassa 3. (Skansi, 2018)



Kuva 3. Kokonainen K-means algoritmin kierto yhdellä sentroidilla. (Mukaillen Skansi, 2018)

Kuvassa 3 esitetyssä määrittäsvaiheessa samanlaiset datapisteet osoitetaan lähimpään satunnaiseen paikkaan asetettuun sentroidiin niiden euklidisen etäisyyden mukaan. Määrittäsvaiheen jälkeen sentroidi siirretään siihen pisteeseen, joka minimoi sentroidiin kuuluvien datapisteiden euklidisen summan. Tätä vaihetta kutsutaan minimointivaiheeksi, joka on myös viimeinen osa algoritmin kiertoa. Minimointivaiheen jälkeen algoritmi aloittaa uuden kierron. Uudessa kierrossa sentroidin paikka pysyy muuttumattomana, mutta se määrittää itselleen uudet datapisteet euklidisen etäisyyden mukaan. Algoritmi toistaa kiertoa niin kauan kunnes halutut klusterit ovat muodostettu. (Skansi, 2018)

Cheong *et al.*, (2017) tutkivat klusterointiin perustuvaa portfolion optimointijärjestelmää, joka käyttää hyödyksi sijoittajien informaatioon perustuvaa geneettistä algoritmia. Heidän tutkimuksensa tavoitteenaan oli selvittää, pystyykö järjestelmää hyödyntämään aktiiviseen salkunhoitoon, ja auttaako järjestelmä suoriutumaan paremmin salkunhoidossa Korean osakemarkkinoilla. He muodostivat ensin klusterointi analyysillä osakeportfoliot sijoittajien eri informaatioista, jonka jälkeen he käyttivät geneettistä algoritmia saadakseen optimaaliset osakepainot portfolioihin. Heidän tutkimustuloksiensa mukaan Sharpen luvulla painotetut ja puolen vuoden algoritmin harjoittelujaksolla muodostetut portfoliot, jotka hyödynsivät institutionaalisen sijoittajan informaatiota, suoriutuivat parhaiten vuosien 2007–2014 välisellä tarkasteluperiodilla. Aiemmin mainituilla parametreilla muodostetut salkut tuottivat vuosittain keskimäärin 24,12 % Sharpen luvun ollessa 0,7 yksikköä. Sharpen luvulla painotetut salkut tuottivat huomattavan suuren riskillä korjatun ylituoton verrattuna markkinaindeksi KOSPI 200, jonka tuotto oli tarkasteluperiodilla vain 2,9 %. (Cheong *et al.*, 2017)

Cheong *et al.*, (2017) tuloksia K-means klusteroinnin suhteen tukee Nandan, Mahantyn ja Tiwarin (2010) tekemä tutkimus. Heidän tutkimuksensa keskittyi portfolion hallintaan ja sen hajauttamiseen. Heidän tavoitteensa oli luoda Markowitz'in teorian mukainen portfolio klusterointi menetelmällä. Tutkimuksessaan he luokittelivat Bombayn pörssin osakkeet aikaisempien tuottojen ja niistä muodostettujen tunnuslukujen perusteella. Osakkeet luokiteltiin klustereihin käyttäen kolmea eri menetelmää, K-means, SOM:ia ja Fuzzy C-means.

Heidän tutkimustuloksiansa mukaan K-means klusterointi luo kompakteimpia klustereita luokkienvälisen inertian valossa. (Nanda *et al*, 2010).

2.5 Jensenin alfa

Jensenin alfa on Jensenin vuonna 1968 suurelle yleisölle esittämä portfolion riskikorjatun suoriutumiskyvyn mittari (Jensen, 1968). Vuosien saatossa siitä tuli yksi rahoitusmaailman tunnetuimmista ylituottoa kuvaavista mittareista, erityisesti erilaisten rahastojen kohdalla. Jensenin esittämä mittari liittyy hyvin läheisesti toisessa alakappaleessa esitettyyn CAP-malliin, jota Jensen käytti pohjana oman mittarin luomisessa. (Jensen, 1968; Sharpe, 1964) Jensenin luomalla Jensenin alfalla pyritään kuvaamaan, kuinka portfolion keskimääräinen tuotto on suoriutunut betalla mitatun CAP-mallin tuottoon nähden. (Jensen, 1968) Jensenin alfaa voidaan kuvata seuraavalla kaavalla:

$$\alpha_i = (r_i - r_f) - \beta_i(r_m - r_f) \quad (3)$$

Jossa α_i edustaa alfaa, r_i kuvaa portfolion tuottoa ja r_f riskitöntä tuottoa. Portfolion systemaattista riskiä yhtälössä kuvaa β_i ja markkinan tuottoa r_m . (Jensen, 1968) Yhtälössä alfan α_i antama tulos kertoo, kuinka valittu portfolio on suoriutunut riskikorjattuun tuottoon nähden. Jos alfa saa arvokseen yli nollan, on portfolio tuotto ylittänyt riskikorjatun tuoton. Tällöin tuotto ylittää aiemmin mainitun arvopaperimarkkinasuoran eli sijoitus on mittarin valossa kannattava. Vastaavasti, jos alfa saa arvokseen pienemmän kuin nolla, on salkun tuotto huonompi kuin riskikorjattu tuotto. Negatiivisilla arvoilla tuotto asettuu arvopaperimarkkina suoran alapuolelle, jolloin sijoitus ei ole kannattava. Alfa voi saada arvokseen myös nollan. Tällöin portfolion tuotto ja riskikorjattu tuotto ovat olleen yhtä suuret, joten tuotto asettuu arvopaperimarkkinasuoran tangentille. (Kallunki, Martikainen & Niemelä, 2019)

2.6 Sharpen luku

Sharpe esitti vuonna 1966 reward-to-variability -suhteen, joka tunnetaan nykyisin Sharpen lukuna. Se edellisessä alaluvussa mainitun Jensenin alfan kanssa toinen rahoitusmaailmaan vakiintunut portfolion suoriutumisen mittari. (Horowitz, 1966) Sharpen luku on vakiintumisesta huolimatta saanut kritiikkiä sen jäykkyydestä, sekä tarkkuudesta, mutta silti mittari on hyvin suosittu esimerkiksi rahastojen vertailussa. (Homm & Pigorsch, 2012) Sharpen luvulla pyritään usein määrittämään ja vertailemaan kuinka hyvin rahasto on suoriutunut koska se ottaa huomioon portfolion riskin. Sharpen luku voidaan esittää seuraavalla kaavalla:

$$\text{Sharpen luku} = \frac{r_i - r_f}{\sigma_i} \quad (4)$$

Jossa kaavan osoittajassa oleva r_i kuvaa portfolion tuottoa ja r_f edustaa riskitöntä tuottoa. Kaavan nimittäjässä oleva σ_i edustaa portfolion keskihajontaa eli volatilitteettiä, joka toimii riskin mittarina. (Sharpe, 1966) Sharpen luku pyrkii kertomaan, kuinka paljon portfolion tuottoon nähden sijoittaja on joutunut kantamaan suhteessa sen kokonaisriskiin. Tällöin, mitä suuremman arvon Sharpen luku saa, sitä kannattavampaa on riskinottaminen, koska tuoton suhde riskiin on parempi. Vastaavasti taas, jos Sharpen luku saa negatiivisen arvon, ei riskin ottaminen ole kannattavaa, koska tuotto on riskitöntä korkoa huonompaa. (Sharpe, 1966)

3 Aineisto ja tutkimusmenetelmät

Tämä tutkielman pääluke koostuu kahdesta kokonaisuudesta, aineistosta ja tutkimusmenetelmistä. Näitä kokonaisuuksia käsitellään kahden alaluvun avulla. Pääluvun ensimmäisessä alaluvussa esitellään tutkielmassa käytetty aineisto. Alaluvun tarkoituksena on kertoa, miten aineisto on syntynyt, sekä mihin tutkielman portfolioita verrataan. Pääluvun toinen alaluku keskittyy tutkielmassa käytettyihin tutkimusmenetelmiin ja portfolioiden muodostamiseen. Samassa kappaleessa käydään myös läpi, miten tutkielmassa käytetyt portfolioit ovat muodostettu.

3.1 Tutkielman aineisto

Tutkielman aineisto koostuu aikasarjadatasta, joka on haettu Thomson Reuters Eikon Databaasista. Aikasarjadata sisältää FTSE-100 indeksissä olevien osakkeiden ensimmäisen päivittäisen kaupankäynti hinnan vuosien 2010–2020 väliltä. Aikasarjadataan kuuluu myös FTSE-100 indeksissä olevien yhtiöiden osakekohtainen tulos, - kirjanpitoarvo, -kassavirta, sekä -osinko tutkielman tarkasteluperiodilta. Tämän aikasarjadataan avulla aineistosta on laskettu Microsoft Exceliä käyttäen P/E, P/B ja P/CF tunnusluvut jakamalla osakkeen vuoden ensimmäisellä noteeraushinnalla jokaisen yhtiön tulos, omaisuuden kirjanpitoarvo, sekä kassavirta. Näitä laskettuja tunnuslukuja käytetään arvoportfolioiden muodostamiseen. Tutkielman aineistosta puuttuu satunnaisten yhtiöiden dataa, johtuen Thomson Reuters databaasin tietokannasta. Nämä puuttuvat yhtiöt ovat poistettu aineistosta, eikä niille ole kerätty dataa tai laskettu tunnuslukuja.

Vertailukohdaksi tutkielmaan on valittu FTSE 100-Total Return indeksi, koska se ottaa huomioon hinnan kehityksen lisäksi osingot. Saatujen portfolioiden kokonais- ja riskikorjattuja tuottoja verrataan tutkielmassa käytettyyn vertailukohtaan, jonka tuottoa esitetään kuvassa 4.



Kuva 4. FTSE 100 Total Return indeksin tuotto ajalta 2010–2020

Kuvassa 4 olevan FTSE 100- Total Return indeksin vuotuinen tuotto vuosien 2010–2020 välillä on ollut 5,51 % ja sen kumulatiivinen tuotto samalla tarkasteluperiodilla on ollut 80,42 %. Laskelmissa ei ole otettu huomioon inflaatiota tai muita kustannuksia, kuten veroja tai transaktiokustannuksia. Sama oletus pätee muihin tutkielmassa tehtäviin laskemiin.

Tutkielman aineistoa valmisteltaessa on otettu huomioon Lontoon pörssissä oleva käytäntö, jonka mukaan osakkeiden hinnat ilmoitetaan sterlingeissä eikä punnissa. Tutkielmassa oletetaan, että yhtiöiden maksamat osingot sijoitetaan takaisin saman yhtiön osakkeeseen, joka on maksanut osingon. Lisäksi laskelmissa on otettu huomioon osakkeiden jakautumiset (engl. Split). Tutkielmassa riskittömänä korkona Jensenin Alfa, betan ja ylituoton määrittämisessä käytetään U.K 10 Year Gilttiä, joka vastaa Ison-Britannian kymmenen vuoden valtionvelkakirjaa. Tutkielman tarkoituksena on tutkia Lontoon pörssin osakkeiden suorittumista, joten valittu riskitön korko soveltuu siihen erinomaisesti.

3.2 Tutkimusmenetelmät ja portfolioiden muodostaminen

Tutkielman empiirinen osio keskittyy portfolioiden muodostamiseen K-means algoritmin avulla. Tutkielmassa käytetään Issakaisen (2022) suunnittelemaa ja luomaa algoritmiä, johon aineistoa luodessa lasketut tunnusluvut syötetään. Luotu algoritmi poistaa

automaattisesti yhtiöt, joiden P/E ja P/CF tunnusluvut ovat negatiiviset. Kuten edellisessä alakappaleessa on mainittu, aineistossa olevat tunnusluvut ovat laskettu jakamalla jokaisen vuoden ensimmäinen osakkeen noteerauspäivän hinta samana vuonna koko kalenterivuodelta mitatulla osakekohtainen tuloksella, - kirjanpitoarvolla ja – liiketoiminnan kassavirralla. Aineistosta lasketut tunnusluvut ovat laskettu vain kalenterivuodelle, joten ne eivät ota huomioon yhtiöiden Q4 tuloksia, jos niiden julkistamisajankohta ylittää kalenterivuoden viimeisen päivän.

Tunnuslukujen laskemisen ja aineiston luonnin jälkeen tunnusluvut syötetään K-means algoritmiin, joka luo tutkielmassa käytetyt portfoliot määritetyn klusteri määrän mukaan. Oletuksena portfolioiden muodostamisessa on, että samassa portfolioissa olevilla osakkeilla on yhtä suuret painoarvot ja ne pysyvät muuttumattomina koko prosessin ajan. Algoritmi poistaa automaattisesti aineistosta poikkeamat (engl. Outlier), jolloin keskiarvot eivät altistu voimakkailla vinoumille. Aineistossa poikkeamana pidetään sellaista osaketta, jonka P/E on suurempi kuin 75 tai P/B on suurempi kuin 10 tai P/CF suurempi kuin 50. Poikkeamien poistamisella voidaan varmistua siitä, että klusteroinnilla saadaan pienin mahdollinen euklidinen etäisyys datapisteiden välillä. (Can *et al.*, 2017) Näin voidaan varmistua siitä, että portfolioihin saadaan vain laadukkaita arvoyhtiöitä. Algoritmi ottaa syötteenä jokaisen tutkielman tarkasteluperiodilla olevan vuoden aineiston, jossa on vain positiivisia P/E, P/B ja P/CF lukuja. Jokaisen vuoden aineisto ajetaan samalla klustereiden määrällä algoritmin läpi, joka tuottaa yhdellä klusterimäärällä aineiston koko tarkasteluperiodille. Tästä aineistosta muodostetaan yksi tarkasteltava arvoportfolio, joka sisältää matalimmat tunnusluvut tai vähintään kaksi kolmesta. Sama toimenpide tehdään vuosittain, jotta indeksiin tulevat uudet osakkeet saadaan sisällytettyä portfolioihin. Portfolioiden muodostamisessa ei huomioida sellaisia osakkeita, jotka ovat poistettu indeksistä vuoden ensimmäisen päivän jälkeen.

Portfolioiden tuottojen laskennassa on otettu huomioon maksetut osingot, koska ne ovat myös sisällytetty vertailukohdan tuottoon. Aineisto syötetään algoritmiin eri klustereiden määrällä, jotta voidaan selvittää, millä klustereiden määrällä sijoittaja pystyy saamaan parhaimman tuoton ja riskikorjatun tuoton. Tuottojen vertailulla pyritään selvittämään, onko arvoanomaliaa esiintynyt tarkasteluperiodilla.

4 Tutkimustulokset

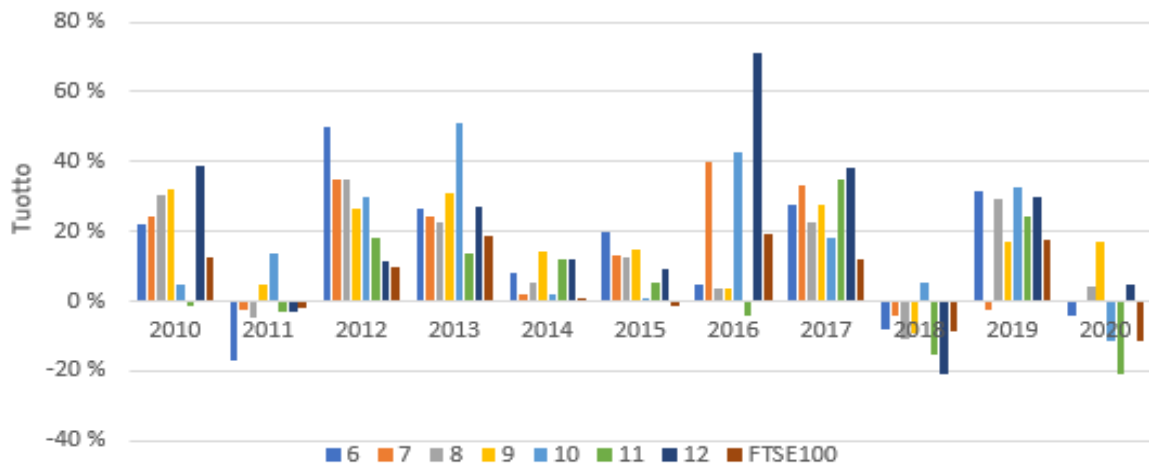
Tässä tutkielman pääkappaleessa esitellään tutkimuksessa saatuja tuloksia. Kappaleessa pyritään esittämään, miten K-means klusterointimenetelmällä muodostetut portfoliot ovat suoriutuneet keskinäisessä vertailussa, sekä vertailukohtaan nähden tutkielmassa käytetyllä tarkasteluperiodilla. Lisäksi pääluvussa esitetään myös millaisia riskikorjattuja tuottoja portfoliot ovat saavuttaneet. Riskikorjattujen tuottojen perusteella valitaan paras klustereiden määrä, jota sijoittajan olisi kannattanut käyttää Lontoon pörssissä vuosien 2010–2020 välillä. Pääluvussa pyritään myös antamaan alustavasti vastamaan, onko Lontoon pörssissä esiintynyt arvoanomaliaa tutkimustulosten valossa.

4.1 Klusterointimenetelmällä luotujen portfolioiden suoriutuminen

Klusterointimenetelmällä luodut portfoliot suoriutuivat erinomaisesti vuosituottojen valossa vuosien 2010–2020 välillä, kuten kuvasta 5 voidaan nähdä. Kuvassa 5 havainnollistetaan, millaisia vuosituottoja eri klusterimäärillä luodut portfoliot ovat tuottaneet tarkasteluperiodilla. Vuosituottojen kohdalla suurta huomiota herättää tuottojen suuret vaihteluvälit, joka tukee tuottojen satunnaisuutta. Tutkielman aineiston perusteella suurin vaihteluväli on 12 klusterilla luodulla portfolioilla. Tämän portfolion vaihteluväli on 92 %, joka koostuu vuoden 2018 21 % tappiosta, sekä vuoden 2016 71 % tuotosta. Vuosi 2018 oli keskiarvallisesti tutkielmassa luotujen portfolioiden huonoin vuosi tuoton kannalta, joka voidaan huomata kuvasta 6. Kuvassa 6 esitetään jokaisen eri klusterimäärällä luotujen arvoportfolioiden keskimääräistä tuottoa ja sitä, miten arvoportfolioiden keskimääräinen tuotto on suoriutunut vertailukohtaan verrattuna. Tuottojen suuret vaihteluvälit selittyvät pääsääntöisesti makrotaloudellisilla ilmiöillä, jotka ovat vaikuttaneet osakkeiden tuottoon ja niiden arvostuksien tasoon.

Vuoden 2011 osakkeiden heikkoon suoriutumiseen vaikutti USA:n osakemarkkinoilla tapahtunut ”musta maanantai”, joka oli seurausta Standard and Poor’sin päätöksestä laskea USA:n luottoluokitusta. Samana vuonna Euroopassa oli meneillään velkakriisi, joka vaikutti

huomattavasti osakkeiden tuottoon sijoittajien pessimistisen taloudellisen luottamuksen takia. Vastaavasti vuonna 2018 Lontoon osakemarkkinoiden tuottoa laski epävarmuudet Brexitistä, sekä USA ja Kiinan välisestä kauppasodasta. Epävarmuudet vaikuttivat eniten suuriin iso-britannialaisiin monikansallisiin yrityksiin, joista FTSE 100- indeksi koostuu. FTSE 100 indeksi laski vuonna 2018 12,5 %¹. Vuoden 2020 heikon tuoton syynä oli aggressiivinen koronaviruksen leviäminen ja siitä johtuneet sulkutilat. Osakemarkkinoiden globaalisuudesta johtuen ympäri maailmaa tapahtuvilla makroekonomisilla tapahtumilla on vaikutus myös Lontoon pörssiin.

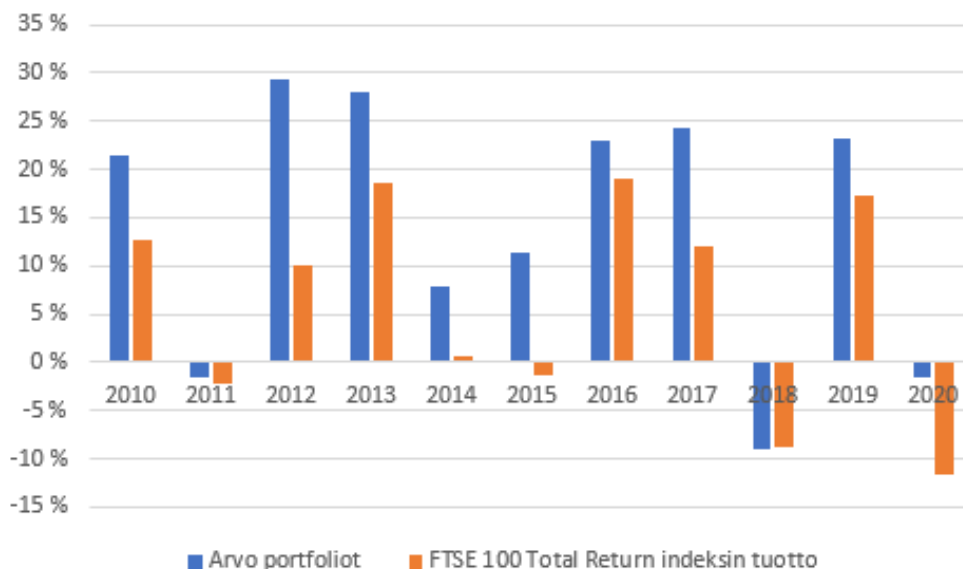


Kuva 5. Klusteroitujen arvoportfolioiden ja FTSE 100- indeksin vuosituotot vuosien 2010–2020 välillä.

Kuvan 5 perusteella arvoportfolioiden vuosituotot näyttävät satunnaisilta, eivätkä niiden jakaumat osoita jakautumista kumpaakaan klusterimäärän ääripäähän. Vaikka vuodet 2011, 2018 ja 2020 olivat huonoja osakkeille ja indeksille tuottojen valossa. Kuvasta 5 voidaan kuitenkin nähdä, että kymmenellä klusterilla luotu portfolio on suoriutunut paremmin kuin tutkielmassa käytetty vertailukohta. Kymmenellä klusterilla luotu portfolio tuotti 14 % vuonna 2011, 5 % vuonna 2018 ja -11 % vuonna 2020. Kun vastaavasti vertailukohdan tuotto vuonna 2011 oli -2 %, -9 % vuonna 2018 ja -12 % vuonna 2020. Tutkielman aineistosta muodostettujen portfolioiden suurin tuotto oli kahdellatoista klusterilla muodostetulla

¹ Tässä tapauksessa indeksillä tarkoitetaan perinteistä FTSE 100 indeksiä, joka ei ota huomioon osinkoja. FTSE 100 Total Return indeksi, joka ottaa huomioon osingot, laski 9 % vuonna 2018.

arvoportfoliolla vuonna 2016, joka oli 70,94 %. Kuvasta 6 voidaan huomata, että tutkielman aineiston kannalta kaikkina vuosina paitsi vuonna 2018, eri klusterimäärillä muodostettujen arvoportfolioiden keskimääräinen vuosituotto on ollut huomattavasti korkeampi kuin vertailukohdalla. Vuoden 2018 kohdalla keskiarvoa alentaa 12 klusterimäärällä luotu arvoportfolio, jonka tappio oli kyseisenä vuonna 20,5 %. Vastaavasti arvoportfoliot ovat keskimääräisesti suoriutuneet huomattavasti paremmin silloin, kun vertailuindeksin vuosituotto on ollut negatiivinen, kuten vuonna 2015 ja 2020. Syynä tähän voi olla sijoittajien käyttäytyminen epävarmoina aikoina, jolloin sijoittajat preferoivat voitollisia ja taloudellisesti stabiileja arvoyrityksiä kasvuyritysten sijaan. Kuvasta 6 voidaan huomata, että tutkielman aineiston kannalta kaikilla klusterimäärillä mitattuna vuodet 2012, 2013 ja 2017 olivat keskiarvallisesti parhaimmat tuottojen osalta.



Kuva 6. Klusteroitujen arvoportfolioiden vuosittainen keskiarvotuotto verrattuna FTSE 100 Total Return indeksin tuottoon

Taulukosta 1, joka esittää arvoportfolioiden tuottoja eri klusterimäärillä, voidaan kuitenkin huomata, että aineisto vinoutuu tuottojen kannalta oikealle. Tämän tutkimuksen tulosten valossa K-means algoritmi on pystynyt havaitsemaan parhaita arvo-osakkeiksi luokiteltuja osakkeita sitä paremmin, mitä suurempi klustereiden määrä on. Mitä suurempi klustereiden

määrä, sitä enemmän portfolion sisältö on keskittynyt arvo-osakkeisiin, joiden tunnusluvut ovat optimaalisella tasolla.

Taulukko 1. Arvoportfolioiden ja vertailukohtan tuotto ajalta 2010–2020.

Klustereiden määrä	6	7	8	9	10	11	12	FTSE100
Kumulatiivinen tuotto	219,2 %	228,70 %	186,08 %	274,1 %	381,32 %	71,51 %	342,9 %	80,42 %
Annualisoitu tuotto	11,13 %	11,42 %	10,03 %	12,74 %	15,36 %	5,03 %	14,10 %	5,51 %

Taulukosta 1 voidaan huomata, että tässä tutkielmassa saavutetut tulokset ovat aiemman arvo-osakkeita sekä arvoanomaliaa tutkivan kirjallisuuden kanssa samankaltaisia tuottojen suhteen² (Capaul *et al.*, 1993; Fama ja French, 1998), pois lukien 11:sta klusterilla muodostettu arvoportfolio. 11:sta klusterimäärällä muodostettu arvoportfolio saavutti vain 5,03 %:in annualisoidun tuoton vuosien 2010–2020 välillä, kun taas tutkielman vertailukohta saavutti 5,51 %:in annualisoidun tuoton. Kaikki muut tutkielmaan valittujen klusterimäärien tuotot voittivat tutkielmassa käytetyn vertailukohtan. Yksi selittävä tekijä 11:sta klusterilla muodostetun arvoportfolion huonoon tuottoon on vuosien 2018 ja 2020 vertailukohtaa suurempi tappio. Vertailukohtaa suurempi tappio ja 11 klusterin heikko suoriutuminen on seurausta yksittäisistä osakkeista, kuten British American Tobacco (-47,20 %), WPP (-32,44 %), Schroders (-27,54 %), ITV (-19,80 %), British Land (-18,63 %) ja Lloyds Banking Group (-41,70 %). Tässä tilanteessa K-means algoritmi on silti valinnut tunnuslukujen perusteella arvo-osakkeita, mutta niiden suoriutuminen on ollut tappiollista ja niiden beta on ollut suuri ottaen huomioon kyseisinä vuosina vallinneen epävarmuuden. Toinen huomio taulukossa 1 on kahdeksalla klusterilla luotu arvoportfolio. Tämä portfolio on poikennut tuottojen kasvusta, mutta sen tuotto silti ollut suurempi kuin tutkielmassa käytetyn vertailukohtan.

Taulukon 1 perusteella 9, 10 ja 12 klusterilla muodostetut arvoportfoliot ovat suoriutuneet parhaiten raakatuottojen perusteella. Kun näitä kolmea eri klusterimäärää tutkii, voi taulukosta huomata, että tuottojen valossa 10 klusterilla luotu arvoportfolio on tuottanut

² Aiemmassa kirjallisuudessa (Capaul *et al.*, 1993; Fama ja French, 1998) arvo-osakkeista muodostetut portfoliot ovat voittaneet tutkielmissa vertailukohtana käytetyn indeksin, kuten tässäkin tutkielmassa arvoportfolioiden tuotto on huomattavasti korkeampi kuin vertailukohtan.

parhaiten. Tällä menetelmällä arvoportfolion luoden, arvosijoittaja olisi voinut saada 381,32 %:in kumulatiivisen tuoton verrattuna vertailukohtaan 80,42 %:iin. Tuottojen erot toisiinsa nähden ovat kuitenkin minimaaliset, pois lukien 11:sta klusterilla saadut tuotot. Toinen syy 11:sta klusterilla saatujen tuottojen heikkouteen ovat arvoportfoliossa olevat yksittäiset osakkeet. Kuvasta 5 voidaan huomata, että 11:sta klusterilla muodostetun arvoportfolion tuotto on ollut vuoteen 2018 kohtalaista. Vuoden 2018 15,03 % tappion jälkeen salkun tuotto palasi erinomaiseksi, mutta laski 21,16 % vuonna 2020. Tappiollisilla osakkeilla oli suuri vaikutus portfolion raakatuottoon, koska osakkeiden kokonaismäärä portfoliossa oli vain 12.

Tässä tutkielmassa saadut tulokset K-means klusteroinnin suoriutumisen portfolion luomisen tukena ovat linjassa aikaisempien tutkimusten kanssa. (Cheong *et al.*, 2017; Mahanty *et al.*, 2010). Aikaisempien tutkimustulosten mukaan K-means klusterointimenetelmä on pysynyt tekemään eri aikaväleillä ja eri osakemarkkinoilla ylituottoja suhteessa markkinaan. Tämän tutkielman tulos osoittaa aikaisempien tutkimusten tavoin, että klusterointimenetelmä on tehokas väline portfolion luomisen tukena. Näiden havaintojen perusteella voidaan todeta, että tutkielman aineistolla ja valitulla aikavälillä K-means klusteroinnilla olisi voinut saada ylituottoja Lontoon osakemarkkinoilta luomalla arvoportfolion 6, 7, 8, 10, 11 tai 12 klusterilla. Tuloksia vertailtaessa on tärkeä muistaa, että tässä tutkielmassa on käytetty parametreinä portfolioiden muodostamisessa vain taloudellisia tunnuslukuja, joista on suodattettu pois poikkeamat (engl. Outlier), sekä negatiiviset arvot. Tästä syystä on hyvä tutkia portfolioiden riskikorjattua tuottoa, joita käsitellään seuraavassa alakappaleessa.

4.2 Arvoportfolioiden riskikorjattu suoriutuminen

Edellisessä aluvuossa esiteltyjen raakatuottojen perusteella klusterointimenetelmän avulla luoduilla arvoportfolioilla pystyttiin saavuttamaan erinomaisia tuottoja verrattuna tutkielmassa käytettyyn vertailukohtaan. Tässä tutkimustuloksien aluvuossa käsitellään klusterointimenetelmällä luotujen arvoportfolioiden tuottamia riskikorjattuja tuloksia. Taulukossa 2 esitellään eri klusterimäärillä luotujen arvoportfolioiden kokoa, sekä portfolioiden suoriutumismittareiden tuottamia tuloksia.

Taulukko 2. Eri klusterimäärien suoriutumismittarit ja portfolioissa olevien osakkeiden määrä

Klusterimäärä	6	7	8	9	10	11	12
Jensenin Alfa	8,06 %	8,70 %	8,00 %	12,23 %	9,04 %	-0,11 %	10,02 %
Sharpen luku	0,70	0,83	0,84	1,24	0,85	-0,26	0,78
Osakkeiden määrä	31	23	18	15	14	12	11

Suoriutumismittareiden tulkinnassa yhteistä on se, että mitä suuremman arvon ne saavat, sitä paremmin ne ovat suoriutuneet suhteessa riskiin. Taulukon 2 perusteella sijoittaja olisi saanut tutkielman tarkasteluperiodilla parhaan riskikorjatun tuoton muodostamalla arvoportfolion yhdeksällä klusterilla. Yhdeksällä klusterilla muodostetun arvoportfolion Sharpen luvuksi muodostui 1,24 yksikköä, joka on huomattavasti muita klusterimääriä suurempi portfolion matalan tuottojen keskihajonnan ansiosta. Seuraavaksi parhaan riskikorjatun tuoton sijoittaja olisi saanut 10 klusterilla muodostetulla arvoportfoliolla, jonka Sharpen luku oli vain 0,85 yksikköä. Taulukon 2 mukaan yhdeksällä klusterilla muodostettu arvoportfolio oli myös Jensenin alfalla mitattuna paras. Yhdeksällä klusterilla muodostetun arvoportfolion Jensenin alfa oli 12,23 %. Portfolio ylitti huomattavalla marginaalilla riskikorjatun tuoton, joka kertoo siitä, että portfolion beta on ollut aineiston matalin. Toiseksi parhaimman Jensenin alfan arvon sain 12 klusterilla muodostettu arvoportfolio, jonka Jensenin alfa oli 10,02 %. Taulukosta 2 voidaan kuitenkin huomata, että vaikka 12 klusterilla muodostetun arvoportfolion Jensenin alfa on erinomainen, on portfolion Sharpen luku vain 0,78 yksikkö. Yksiksi 12:sta klusterilla muodostetun arvoportfolion Sharpen luvun heikkouteen on kyseisen portfolion saavuttamien tuottojen suuri vaihteluväli, joka on aineiston mukaan kasvattaa portfolion tuottojen keskihajontaa eli sen kokonaisriskiä huomattavasti.

Taulukosta 2 voidaan huomata, että kokonaisriskiä mittaavan Sharpen luvun kehitys on ollut nousujohteista, vaikka Sharpen luku on laskenut verratessa 9:llä ja 10 klusterilla luotuja arvoportfolioita. Kuitenkin 10 klusterilla luodun arvoportfolion Sharpen luku on suurempi kuin 6:lla, 7:llä ja 8 klusterilla luotujen portfolioiden. Tämän havainnon mukaan arvoportfolioissa olevien osakkeiden määrä on luonut tarpeeksi hajautusta eri klusterimäärillä,

vaikka osakkeiden määrä portfoliossa on pienentynyt. Riittävä hajautus ja siitä saatava hyöty voidaan huomata siitä, että portfolioiden kokonaisriski on laskenut 10 klusterimäärään saakka, mutta Sharpen luku on kääntynyt laskuun. Sharpea lukua ja Jensenin alfaa tulkittaessa on kuitenkin hyvä muistaa, että portfolioissa olevien osakkeiden painoarvot ovat yhtä suuret, eikä näin ollen portfolioita ole painotettu Markowitzin portfolioiteorian mukaan riskiprofiiliin sopivaksi. Lisäksi osakkeiden tuottojen keskinäisiin korrelaatioihin ei ole kiinnitetty huomiota portfolioiden luomisen yhteydessä, joten portfolioit eivät ole tehokkaan rintaman mukaisia, ja ne sisältävät vain yhtä omaisuusluokkaa.

Taulukon 2 esittämien tulosten perusteella kaikissa muissa paitsi 11 klusterilla muodostetulla portfolioilla riskinottaminen olisi ollut kannattavaa tarkasteluperiodilla, koska näiden portfolioiden Sharpen luku oli ollut positiivinen. 11 klusterilla muodostetun salkun kohdalla sijoittajan olisi kannattanut sijoittaa tutkielmassa käytettyyn vertailukohtaan, jolloin sijoittaja olisi saanut paremman tuoton, eikä olisi joutunut kantamaan riskiä sijoituksistaan kuin systemaattisen riskin eli markkinariskin verran.

Muodostettujen arvoportfolioiden keskimääräinen Sharpen luku aineiston perusteella oli 0,69 yksikköä. Vastaavasti arvoportfolioiden keskimääräinen Jensenin alfa oli 7,99 %. Taulukosta 2 voidaan huomata, että suoriutumismittareiden keskinäinen vaihteluväli eri klusterimäärillä oli huomattavan suuri. Eri klusterimäärillä muodostettujen portfolioiden Sharpen luvun vaihteluväli oli -0,26–1,24 ja kokonaisvaihteluväli oli 0,98 yksikköä. Vastaavasti muodostettujen arvoportfolioiden Jensenin alfan vaihteluväli oli -0,11 %-12,23 % ja kokonaisvaihteluväli oli 12,12 %.

4.3 Tulosten pohdinta

Klusterointimenetelmällä luodut arvoportfolioit suoriutuivat erinomaisesti riskikorjattujen tuottojen vertailussa. Arvoportfolioit saavuttivat riskikorjattuja ylituottoja, pois lukien se

portfolio, joka hävisi tutkielman vertailukohdalle. Tämä osoittaa, että K-means klusterointimenetelmä pystyy valitsemaan laadukkaita osakkeita tunnuslukujen avulla. Havaintoa tukee Nanda *et al.*, (2010) tekemä tutkimus klusterointimenetelmän tehokkuudesta. Tärkeä huomio tutkimuksessa on se, että arvoportfolioiden riskikorjattua tuottoa pystyttiin kasvattamaan huomattavasti, kun klusterointialgoritmi suodatti pois yritykset, joiden tunnusluvut olivat negatiiviset. Kun yhtiöt suodatettiin aineistosta pois, sillä pystyttiin vaikuttamaan arvoportfolioiden kokonaisriskiin. Tämä pystyttiin huomaamaan siitä, että sijoittaja pystyi saamaan suurempia tuottoja, vaikka portfolioissa olevien osakkeiden määrä laski ja portfolion Sharpen luku kasvoi.

Tutkielman aineiston perusteella vaikuttaa siltä, että arvoportfolioiden saavutetut korkeat tuotot johtuvat pääsääntöisesti Dissanaike ja Lim (2010) mainitsemaasta paluu keskiarvoon ilmiöstä. Paluu keskiarvoon ilmiö on aiemman kirjallisuuden mukaan yksi merkittävimmistä syistä arvoanomalioiden syntymiseen (Fama ja French, 2007; Lakonishok, Shleifer ja Vishny, 1994). Arvoportfoliot luonti menetelmän mukaan niihin valikoitui sellaisia osakkeita, jotka olivat tunnuslukujen valossa aliarvostettuja. Kun aliarvostettuja osakkeita pidettiin portfolioissa, niistä suurimman osan arvostus palautui takaisin oikeaan arvoonsa. Tämän seurauksena osakkeiden tunnuslukujen P eli hintakomponentti palasi takaisin keskiarvoonsa, eikä tunnusluvun nimittäjä kasvanut yhtä nopeasti.

5 Johtopäätökset ja yhteenveto

Tutkielmassa tavoitteena oli selvittää kahden kokonaisuuden avulla, miten K-means algoritmiin syötettävien klustereiden määrä vaikuttaa portfolioiden kokonaistuottoon, ja onko K-means algoritmeilla pystynyt löytämään arvoanomalian Lontoon osakemarkkinoilta vuosien 2010–2020 välillä. Tutkielman ensimmäinen kokonaisuus muodostui kokonaisuuden kannalta relevanteista teorioista, sekä aiemmasta kirjallisuudesta, joita esiteltiin tutkielman toisessa pääluvussa. Tutkielmassa käsitelty aikaisempi kirjallisuus antoi vahvaa näyttöä siitä, että Lontoon osakemarkkinoilta on löytynyt arvoanomaliaita matalien P/E, P/CF ja P/B tunnuslukujen avulla (Capaul *et al.*, 1993; Fama ja French, 1998; Andersson & Brooks, 2006). Toinen kokonaisuus muodostui empiirisestä osiosta, joka suoritettiin kvantitatiivisena tutkimuksena. Toisen kokonaisuuden tarkoituksena oli tutkia K-means algoritmin toimintaa ja sen avulla muodostettujen arvoportfolioiden suoriutumista Sharpen luvulla, Jensenin alfalla ja raakatuotoilla.

Tutkielman ensimmäisessä tutkimuskysymyksessä pyrittiin selvittämään, *Voidaanko FTSE 100:sta löytynyt arvoanomaliaa vuosien 2010–2020 välillä käyttämällä K-means menetelmää?* Tutkielman tutkimustulosten perusteella voidaan sanoa, että Lontoon pörssistä on voinut löytää arvoanomalian luomalla arvoportfolioita käyttämällä K-means klusterointimenetelmää. Saavutetut tutkimustulokset ovat hyvin samankaltaisia, kuin K-means klusterointimenetelmää osakkeiden valitsemisessa tutkiva kirjallisuus (Cheong *et al.*, 2017; Mahanty *et al.*, 2010). Tutkielman vertailukohtaan voittaneet arvoportfoliot olivat käyttäytyneet aiemman kirjallisuuden tavoin CAP- mallin ja tehokkaiden markkinoiden odotusten vastaisesti. Yksi tutkielman keskeinen havainto oli, että arvoportfolioissa olevat arvo-osakkeet altistuivat paluu keskiarvoilmiöön, jonka seurauksena tuotot olivat huomattavasti suurempia verrattuna tutkielman vertailukohtaan. Kuitenkaan jokaisella tutkielmassa käytetyllä klusterimäärällä ei olisi voinut löytää arvoanomaliaa Lontoon osakemarkkinoilta. Tutkielmassa 11 klusterilla luotu arvoportfolio tuotto oli heikompi kuin vertailukohtaan tuotto, eikä sijoittajan olisi kannattanut ottaa portfolion tuomaa riskiä, koska sen Sharpen luku oli negatiivinen.

Tutkielman toisessa tutkimuskysymyksessä pyrittiin selvittämään, *Kuinka paljon valittu klustereiden kokonaismäärä olisi vaikuttanut kokonaistuottoihin vuosien 2010–2020 aikana?* Arvoportfolioiden tuottamien kokonaisutuottojen annualisoidussa vaihtelussa oli eroja, jotka johtuivat yksittäisten osakkeiden vaikutuksesta portfolion tuottoon. Kumulatiivisena tuottona tarkasteltaessa klustereiden määrä vaikuttaa huomattavasti saatuun kokonaistuottoon. Arvoportfolioiden kumulatiivisten tuottojen vaihteluväli oli 71,51 % - 381,32 % heikoimman tuotolla mitatun portfolion ja parhaan portfolion välillä. Tutkielman tulosten perusteella pystyttiin huomaamaan kokonaistuottojen vinoutuminen oikealle. Tarkoittaen sitä, että, mitä suuremman määrän klusterointialgoritmi ottaa klustereita syötteenä, sitä suuremman tuoton ne tuottivat tässä tutkimuksessa. Klustereiden määrän kasvaessa algoritmi pystyy valitsemaan laadukkaita yhtiöitä, mutta niiden määrä portfoliossa laskee huomattavasti, jolloin portfolio on alttiimpi yhden osakkeen arvonmuutokselle.

Tutkimustulosten mukaan sijoittajan olisi kannattanut muodostaa arvoportfolio 10 klusterilla saadakseen parhaimman annualisoidun tuotot Lontoon osakemarkkinoilta vuosien 2010–2020 välillä. Riskikorjatulla tuotolla mitattuna 10 klusterilla luotu arvoportfolio ei olisi kuitenkaan ollut paras valinta sijoittajalle. Jensenin alfalla ja Sharpen luvulla mitattuna sijoittajan olisi kannattanut luoda arvoportfolio yhdeksällä klusterilla. Yhdeksällä klusterilla luodun arvoportfolion Jensenin alfa oli 12,23 % ja sen Sharpen luku oli 1,24 yksikköä. Kun vastavasti 10 klusterilla luodun arvoportfolion Jensenin alfa oli 9,04 % ja sen Sharpen luku oli 0,85 yksikköä.

Kokonaisuutena tämä tutkielma tarjoaa erinomaisen vaihtoehdon osakkeiden poimimiseen ja arvosijoittamiseen. Tämän tutkielman perusteella yritys, jolle tämä tutkielma tehdään, pystyy saamaan laajan ymmärryksen siitä, miten K-means klusterointimenetelmä toimii osakkeiden valinnassa. Tutkielman keskeisin löytö oli K-means algoritmin kyky löytää laadukkaita yhtiöitä, joilla sijoittaja olisi voinut saada erinomaisia riskikorjattuja tuottoja. Lisäksi tämä tutkimus vahvistaa aikaisempien K-means klusterointimenetelmää osakkeiden valinnassa tutkivien tutkijoiden tuloksia siitä, että K-means on tehokas tapa löytää laadukkaita yhtiöitä. Tutkimus vahvistaa myös aikaisempia arvoanomalia tutkimuksia. Erona aikaisempiin arvoanomalia tutkimuksiin on se, että tässä tutkimuksessa käytettiin yhden

tunnusluvun sijasta kolmea. Uutena tietona tutkimuksessa selvisi, että K-means algoritmilla muodostetuilla portfolioilla on pystynyt saamaan ylituottoja Lontoon osakemarkkinoilta maltillisella riskillä.

5.1 Tutkimustulosten luotettavuuden arviointi ja jatkotutkimus

Tutkimustulosten luotettavuuden arvioinnissa on otettava huomioon suoriutumismittareiden kohtaama kritiikki. Rahoitusmarkkinat ovat muuttuneet huomattavasti siitä lähtien, kun Sharpen luku ja Jensenin alfa on kehitetty kuvaamaan portfolioiden ja osakkeiden riskikorjattua suoriutumista. Gussetin ja Zimmermanin (2014), sekä Goldberg ja Korajczykkin (2010) mukaan suoriutumismittarit ovat jäykkiä, eivätkä ne pysty huomioimaan jatkuvia rahoitusmarkkinoilla tapahtuvia muutoksia.

Suoriutumismittareiden kohtaama kritiikki on pääsääntöisesti kohdistunut niissä olevaan riskikomponenttiin, jota tulkitaan Sharpen luvussa ja Jensenin alfassa Pätäriin (2000, 28–29, 58; Pätäri 2000, 36–37, 42–43) mukaan väärin. Lisäksi Sharpen luvulle ja Jensenin alfalle määritelty suoriutumisasteet ovat hyvin tulkinnanvaraisia ja ne vaihtelevat usein, eikä niille ole määritelty tarkkoja asteita. Pätäri (2000) mainitsee myös tutkimuksessaan, että riskiluvun määrittäminen on usein vinoumille altista. Määrittäminen riippuu paljon siitä, kuinka optimistisellä tai pessimistisellä riskin arvolla suoriutumismittaria tulkitaan, koska sijoittaja pyrkii huomioimaan tulkinnassa rahoitusmaailman muutokset.

Tutkielman reliabiliteettia arvioidessa on hyvä muistaa, että tutkielman osakkeet rajoittuivat vain FTSE 100 indeksin sisältämiin osakkeisiin, joita on vain 100. Tämä edustaa marginaalista osaa koko Lontoon pörssissä olevista osakkeista. Jos tutkielmaan valittu osakkeiden otoskoko olisi suurempi, voisi sillä olla merkittävä vaikutus tutkimustuloksien vaihteluun. Lisäksi tutkielmassa oletetaan, että osakkeiden myynnistä ei koidu veroseuraamuksia, eikä kaupankäynnissä ole transaktiokustannuksia. Jos myynnistä koituvat veroseuraamukset ja transaktiokustannukset otettaisiin huomioon, olisi sillä huomattava vaikutus tutkielman arvoportfolioiden tuottoon.

Tutkielmassa käytetystä klusterointimenetelmästä sijoitusstrategiana olisi hyvä tehdä jatko-tutkimusta. Tämän tutkimuksen tulosten valossa klusterointimenetelmä suoriutui erinomai- sesti, koska klusterointialgorimin tehtävänä oli valita osakkeet, joiden arvostus oli tunnuslu- kujen perusteella alhainen. Erinäisistä syistä, kuten makrotaloudellisista syistä tai sijoittajien ylioptimistisuudesta johtuen yksittäiset osakkeet voivat tulla yliarvostetuiksi. Tästä syystä jatkotutkimuksen tulisi kohdistua siihen, milloin osakkeiden myynti olisi optimaalisinta. Toisin sanoen, kuinka paljon tunnusluvun P komponentti saa nousta, jolloin osake on yliar- vostettu, jonka seurauksena osakkeen hinnan laskun ja siitä koituvan tappion todennäköisyys kasvaa.

Lähteet

- Anderson, K. and Brooks, C. (2006), “The long-term price-earnings ratio”, *Journal of Business Finance and Accounting*, Vol. 33 No. 7–8.
- Basu, S. (1977) Investment Performance of Common Stocks in Relation to Their PriceEarnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance*. Vol. 32, (3) 663–682.
- Basu, S. (1983) The Relationship Between Earning’s Yield, Market Value and Return for NYSE Common Stocks: Future Evidence, *Journal of Financial Economics*, Vol. 12, No. 1, pp. 129–156
- Ball, R., Gerakos, J., Linnainmaa, J.T. and Nikolaev, V. (2020), “Earnings, retained earnings, and book-to-market in the cross section of expected returns”, *Journal of Financial Economics*, North-Holland, Vol. 135 No. 1, pp. 231–254.
- Bodie, Z. et al. (2005) *Investments*. 6th ed. Boston (MA): McGraw-Hill.
- Black, A. J. McMillan, D. G. (2006) Asymmetric risk premium in value and growth stocks. *International Review of Financial Analysis*. Vol. 15, 237–246.
- Capaul, C., Rowley, I., & Sharpe, W. F. (1993). International value and growth stock returns. *Financial Analysts Journal*, 49(1), 27–36.
- Cheong, D., Kim, Y.M., Byun, H.W., Oh, K.J. and Kim, T.Y. (2017), “Using genetic algorithm to support clustering-based portfolio optimization by investor information”, *Applied Soft Computing*, Elsevier, Vol. 61, pp. 593–602.
- Dissanaike, G. and Lim, K.H. (2010), “The sophisticated and the simple: The profitability of contrarian strategies”, *European Financial Management*, Vol. 16 No. 2.
- Elze, G. (2012) Value investor anomaly: return enhancement by portfolio replication – an empiric portfolio strategy analysis. *Central European journal of operations research*. Vol. 20 (4), 633–647.
- Fama, E. F. (1970) Efficient Capital Markets: A Theory and Empirical Work. *Journal of Finance*. Vol. 5 (2), 383–417

- Fama, E.F. (1991). Efficient Capital Markets II. *Journal of Finance*, 46, 5, 1575–617
- Fama, E. F. French, K. R. (1998) Value versus Growth: The International Evidence. *The Journal of Finance*. Vol. 53 (6), 1975–1999.
- Fama, E. F. French, K. R. (2004) The Capital Asset Pricing Model: Theory and Evidence. *The Journal of Economic Perspectives*. Vol. 18 (3) 25–46
- Fama, E. F. French, K. R. (2007) The Anatomy of Value and Growth Stock Returns. *Financial Analyst Journal*. Vol. 63. (6), 44–54
- Forgy, E. W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21(3), 768–769 (1965)
- Frankfurter, G. M. McGoun, E. G. (2001) Anomalies in finance. What are they and what are they good for? *International Review of Financial Analysis*. Vol. 10, 407–429.
- Gan, G. & Ng, M. K.-P. (2017) k-means clustering with outlier removal. *Pattern recognition letters*. [Online] 908–14. Saataavilla: <https://doi.org/10.1016/j.patrec.2017.03.008>
- Graham, B. & Dodd, D. (1934) *Security Analysis*. New York: McGraw-Hill
- Goldberg, C. & Korajczyk, R. A. (2010) *Portfolio risk analysis*. Princeton University Press.
- Gusset, & Zimmermann, H. (2014). Why not use SDF rather than beta models in performance measurement? *Financial Markets and Portfolio Management*, 28(4), 307–336.
- Horowitz, I. (1966) The “Reward-to-Variability” Ratio and Mutual Fund Performance. *The Journal of Business* (Chicago, Ill.), 39 (4), 485–488.
- Homm, U., & Pigorsch, C. (2012) Beyond the Sharpe ratio: An application of the Aumann–Serrano index to performance measurement. *Journal of Banking & Finance*, 36 (8), 2274–2284
- Issakainen, T. (2022) The performance of value investing strategy in Helsinki Stock exchange from 2005 to 2021: A Clustering approach. [Pro Gradu]. Saataavilla: <https://lut-pub.lut.fi/bitstream/handle/10024/163936/Master%27s%20Thesis%2c%20Topi%20Issakainen%2020.4.2022.pdf?sequence=1&isAllowed=y>
- Jensen, M. C (1968) The performance of mutual funds in the period 1945–1964. *The Journal of finance* (New York). Vol. 23 (2), 389–416

- Knüpfer, S. Puttonen, V. (2018) *Moderni Rahoitus*. 10. painos. Helsinki, Alma
- Kallunki, Martikainen, M., & Niemelä, J. E. (2019). *Ammattimainen sijoittaminen* (8., uudistettu painos.). Alma Talent
- Kallunki, J.-P. & Niemelä, J. (2012) *Osakkeen arvonmääritys: onnistunut sijoituspäätös*. Helsinki: Talentum
- Lakonishok, J., Shleifer, A. and Vishny, R.W. (1994), “Contrarian Investment, Extrapolation, and Risk”, *The Journal of Finance*, Vol. 49 No. 5
- Lintner, J. (1965), “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”, *The Review of Economics and Statistics*, Vol. 47 No. 1.
- Lloyd, S. (1982) Least squares quantization in PCM. *IEEE transactions on information theory*. [Verkkoartikkeli] 28 (2), 129–137. Saatavilla: <https://doi.org/10.1109/TIT.1982.1056489>
- Markowitz, H. (1952) Portfolio selection. *The Journal of finance*. Vol. 7 (1), 77–91.
- Markowitz, H. (1999) The Early History of Portfolio Theory: 1600–1960. *Financial Analyst Journal*. 5–16
- Malkiel, B. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17, 59–82.
- Perold, A. F. (2004) The Capital Asset Pricing Model. *The Journal of economic perspectives*. Vol. 18 (3), 3–24
- Pätäri, E. (2000) *Essays on portfolio performance measurement*. Väitöskirja. Lappeenrannan teknillinen korkeakoulu. Lappeenranta
- Sharpe, W.F. (1964), “Capital Asset Prices: A Theory Of Market Equilibrium Under Conditions Of Risk”, *The Journal of Finance*, Vol. 19 No. 3
- Sharpe, W.F. (1966), “Mutual Fund Performance”, *The Journal of Business*, Vol. 39 No. 1, pp. 119–138.
- Stattman, D. (1980), “Book values and stock returns”, *The Chicago MBA: A Journal of Selected Papers*, Vol. 4 No. 1, pp. 25–45.

Skansi, S. (2018), *Introduction to Deep Learning*, Springer International Publishing, Cham

Nicholson, S.F. (1960), “Price-to-earnings ratio”, *Financial Analysts Journal*, *Growing Science*, Vol. 16 No. 4, pp. 43–45.

Nanda, S.R., Mahanty, B. and Tiwari, M.K. (2010), “Clustering Indian stock market data for portfolio management”, *Expert System with Applications*, Vol. 37 No. 12, pp. 87

