

Algoritmiske handelsstrategier i et retailperspektiv

En empirisk vurdering av RSI mean-reversion, faktorrotasjon og multifaktorstrategier

HULC AS

Systematiske handelsstrategier - delrapport

Noah Lekhal Husnes

Student ved Handelshøyskolen BI

August 2026

1. Innledning

1.1 Bakgrunn og motivasjon

Den første boken jeg leste om aksjehandel presenterte Relative Strength Index som et verktøy for å finne oversolgte aksjer som var modne for rekyl. Ideen er tilforlatelig. En mekanisk indikator peker ut kjøpsmuligheter, uten skjønn og uten selskapsanalyse. Den satte i gang et spørsmål som denne rapporten er et forsøk på å besvare, nemlig om kortsiktige prisavvik er målbare og lar seg utnytte systematisk.

HULC AS ble stiftet i januar 2026 som et investeringsselskap med formål å utvikle strukturerte investeringsstrategier forankret i finansiell teori. Selskapets virksomhet er organisert rundt tre komplementære områder: systematisk handel, fundamental analyse, og risikostyring og porteføljekonstruksjon. Denne delrapporten omhandler systematisk handel.

Den teoretiske antakelsen som ligger til grunn for hele prosjektet er at markedet er overveiende effisient på lang sikt, men at kortsiktige avvik og prisfeil oppstår jevnlig og potensielt kan utnyttes systematisk. Denne delrapporten forsøker å teste den siste delen av antakelsen empirisk.

1.2 Forskningsspørsmål og hypotese

Hypotesen ble formulert slik ved oppstarten av arbeidet:

Hypotese 1: Filtrerte RSI-baserte mean reversion-strategier kan gi positiv risikojustert avkastning etter transaksjonskostnader.

Denne hypotesen utgjør utgangspunktet for arbeidet. Når empiriske resultater senere viste at hypotesen ikke kunne bekreftes, ble undersøkelsen utvidet til å omfatte alternative algoritmiske tilnærminger. Denne utvidelsen er ikke et avvik fra den opprinnelige planen, men en logisk konsekvens av forskningsmetoden. Når en hypotese falsifiseres, søker forskningen etter alternative forklaringer eller relaterte hypoteser som kan testes.

Det utvidede forskningsspørsmålet kan formuleres slik:

Finnes det algoritmiske handelsstrategier som, gitt realistiske transaksjonskostnader og skatteforhold for et norsk investeringselskap, kan levere risikojustert avkastning som overgår en passiv referanseportefølje?

Kapitalstørrelsene som brukes gjennom rapporten, 50 000 og 100 000 kroner, er ikke en forenkling. De er en del av spørsmålet. Litteraturen som motiverer strategiene er i hovedsak skrevet på institusjonelle forutsetninger, der ordren er stor nok til at faste kostnader per ordre er uten betydning. Den forutsetningen holder ikke for et lite selskap, og den holder i enda mindre grad for en privatperson. Begge handler i det samme markedet som aktører med kapital, infrastruktur og analyseressurser i en helt annen størrelsesorden. Rapporten tester derfor ikke bare om strategiene har edge, men om edge overlever kostnadene ved å være liten. Det er et annet spørsmål enn litteraturen stiller, og resultatene i kapittel 4 gir et annet svar.

Det samme gjelder datatilgangen. Institusjonelle databaser som Bloomberg, LSEG og CRSP/Compustat er utgangspunktet i litteraturen som motiverer strategiene, og et abonnement på den typen kilde koster mer i året enn kapitalbasen strategiene her er testet med. Tilgangen er derfor ikke et valg som forelå, den er en egenskap ved den aktøren rapporten handler om. Rapporten tester hva et selskap av denne størrelsen faktisk kan måle, med de kildene et selskap av denne størrelsen faktisk kan kjøpe, og det er en del av spørsmålet på samme måte som ordrestørrelsen er det.

1.3 Rapportens struktur

Rapporten er strukturert som følger. Kapittel 2 presenterer det teoretiske rammeverket og en gjennomgang av de strategiene som er identifisert som relevante å teste, basert på en kombinasjon av tradisjonell finansiell teori, akademisk forskning og pragmatiske vurderinger knyttet til retail-investorens praktiske begrensninger. Kapittel 3 beskriver metodikken som er anvendt, inkludert backtestrammeverket, kostnadsmodelleringen og de statistiske målene som benyttes. Kapittel 4 til 6 presenterer resultater, tolkning og konklusjon, samt en diskusjon av implikasjoner for de to øvrige områdene.

1.4 Avgrensninger

Rapporten behandler kun systematiske, regelbaserte strategier. Strategier som krever fundamental analyse av enkeltsselskaper eller overordnet porteføljekonstruksjon og risikostyring hører til de to øvrige områdene, og er ikke en del av denne delrapporten. De vil bli behandlet i separate dokumenter.

Videre er rapporten avgrenset til strategier som er praktisk gjennomførbare for HULC AS som retailaktør. Dette innebærer at strategier som forutsetter institusjonell infrastruktur, lavlatens-tilgang eller markedsdata på nivåer utover hva som er tilgjengelig gjennom Interactive Brokers, ikke er vurdert.

Det er videre viktig å understreke en grunnleggende ressursmessig avgrensning. Som et lite norsk investeringsselskap har HULC AS begrenset tilgang til kostbare datakilder, profesjonell infrastruktur og dedikert utviklingstid. Rapporten dekker derfor det vi har vurdert som de mest representative og relevante overordnede tilnærminger innenfor systematisk handel - mean-reversion, momentum, trendfølging, faktorrotasjon og multifaktorstrategier, men går ikke fullstendig i dybden på hver enkelt tilnærming. Denne avgrensningen påvirker hvor kategoriske konklusjoner vi kan trekke.

2. Teoretisk rammeverk og strategiutvelgelse

2.1 Markedseffisiens-hypotesen

Et naturlig utgangspunkt for enhver vurdering av algoritmiske handelsstrategier er Eugene Famas markedseffisiens-hypotese (Fama, 1970). Hypotesen postulerer at finansmarkeder reflekterer all tilgjengelig informasjon i sine priser, og at det dermed ikke skal være mulig å oppnå systematisk meravkastning gjennom analyse av historiske priser eller offentlig tilgjengelig informasjon. Fama presenterte tre former av hypotesen: svak, semi-sterk og sterk effisiens.

Algoritmiske handelsstrategier basert på tekniske indikatorer som RSI er en direkte test av svak markedseffisiens. Dersom svak effisiens holder, skal historiske kursmønstre ikke kunne brukes til å forutsi fremtidige prisbevegelser, og strategier basert på slike mønstre skal følgelig ikke gi systematisk meravkastning etter kostnader.

Det er imidlertid omfattende empirisk forskning som dokumenterer avvik fra markedseffisiens, særlig på kort sikt. Disse avvikene kalles ofte anomalier i akademisk litteratur, og inkluderer blant annet momentum-effekten, mean-reversion på kort sikt, og forskjellige faktorpremier som value, quality og low volatility. Anomaliene utgjør grunnlaget for at det fortsatt forskes og handles på systematiske strategier til tross for at den teoretiske forventningen er at slike strategier ikke skal fungere.

2.2 Mean-reversion som teoretisk ramme

Mean-reversion er en hypotese om at aktivapriser har en tendens til å vende tilbake mot et statistisk gjennomsnitt over tid. Hypotesen ble systematisk utforsket av blant andre Jegadeesh (1990) og Lehmann (1990), som dokumenterte at aksjer som hadde underprestert på kort sikt (typisk en uke til

en måned) viste statistisk signifikant rekyl i den påfølgende perioden. De Bondt og Thaler (1985) demonstrerte tilsvarende effekter på lengre tidshorisonter.

RSI, utviklet av J. Welles Wilder (1978), er en av de mest brukte tekniske indikatorene for å operasjonalisere mean-reversion. Indikatoren måler styrken i de senere prisbevegelsene og normaliserer denne til en skala fra 0 til 100. Konvensjonelt betraktes verdier under 30 som indikasjon på oversolgt tilstand, og verdier over 70 som indikasjon på overkjøpt tilstand. En enkel mean-reversion-strategi går ut på å kjøpe når RSI faller under 30, med forventning om at prisen vil normalisere seg mot gjennomsnittet.

Til tross for den teoretiske appellen og den intuitive forståelsen, viser nyere forskning at edge-en knyttet til ren RSI mean-reversion har blitt vesentlig svekket i moderne markeder. Dette tilskrives flere faktorer: bredere institusjonell deltakelse, høyere markedseffisiens, og særlig at strategien har blitt så velkjent at den selv påvirker prisdannelsen i de tilfellene hvor den faktisk gir signaler. Mekanismen er dokumentert bredt. McLean og Pontiff (2016) finner at avkastningen til publiserte anomalier faller vesentlig etter at de er publisert.

2.3 Strategier vurdert for testing

Basert på den teoretiske gjennomgangen og praktiske vurderinger ble følgende strategier identifisert som relevante å teste. Disse representerer ulike grader av kompleksitet og er motivert av forskjellige delaspekter ved den overordnede hypotesen. Til sammen dekker de hovedkategoriene av systematiske handelsstrategier som er aktuelle for en retailaktør: mean-reversion, momentum (både tidsserie og kryssektoral), trendfølging og multifaktortilnærminger.

2.3.1 Intraday RSI mean-reversion

Den opprinnelige strategien fra søknaden, basert direkte på Hypotese 1. Strategien identifiserer aksjer som har falt kraftig ved børsåpning på NYSE og som har RSI under 30 på intradag-bars. Posisjoner åpnes med faste prosent-baserte bracket-ordrer (3 % take profit, 1.5 % stop loss) og lukkes ved time-stop etter 90 minutter dersom ingen av brackets utløses.

Strategien representerer den enkleste implementeringen av hypotesen og fungerer som baseline for videre arbeid. Den er motivert av tradisjonell teknisk analyse og er en direkte operasjonalisering av RSI-litteraturen.

2.3.2 Filtrerte intraday-varianter

Når baseline-strategien ikke viste edge, ble fem filtre identifisert i litteraturen som potensielt forbedrende. Disse ble testet både i kombinasjon og hver for seg, for å isolere effekten av hvert enkelt filter:

- Trend-filter: Kun ta entry når aksjekursen ligger over 200-dagers glidende gjennomsnitt. Motivert av forskning som indikerer at mean-reversion fungerer best i aksjer som er i etablert oppadgående trend (Connors og Alvarez, 2009).

- Gapstørrelse filter: Kun ta entry når åpningsfallet er innenfor et bestemt intervall (-8 % til -3 %). Motivert av observasjonen at moderate fall ofte er flow-drevne mens ekstreme fall ofte er fundamentalt drevne (Da, Liu og Schaumburg, 2014).
- Reversal confirmation: Vent til en grønn 5-minutters bar er observert før entry. Motivert av at man da kjøper faktisk reversering snarere enn fortsettelse av nedtrend.
- ATR-basert SL/TP: Erstatt faste prosent-stops med stops basert på Average True Range. Motivert av at volatilitetsjusterte stops gir mer konsistent risikoeksponering på tvers av aksjer.
- Risikobasert posisjonsstørrelse: Beregne antall aksjer slik at dollar-risiko per handel utgjør en fast prosent av porteføljen. Motivert av Kelly-kriteriet (Kelly, 1956) og moderne risikostyring.

2.3.3 Cross-sectional momentum på enkeltaksjer

En klassisk og veldokumentert strategi der aksjer rangeres etter avkastning over en periode (typisk 12 måneder eksklusiv siste måned), og porteføljen holder de aksjene som har rangert høyest. Strategien ble først dokumentert av Jegadeesh og Titman (1993) og har vært gjenstand for omfattende videre forskning.

Strategien ble innledningsvis vurdert som ikke anvendelig for HULC AS. Begrunnelsen var at amerikanske aksjer ligger utenfor fritaksmetoden, slik at hyppig realisering av gevinst medfører en strukturell kostnad på 22 % av realisert gevinst. Vurderingen ledet til at enkeltaksjestrategier ble flyttet til EØS-instrumenter.

Denne vurderingen er senere revidert. Den veide skattekostnaden isolert, uten å holde den opp mot de øvrige strukturelle kostnadene ved den kapitalstørrelsen strategien faktisk skal drives med. Regnestykket er gjennomgått i kapittel 3.1.3. Konklusjonen der er at rangeringen mellom markedene ikke lar seg avgjøre på det grunnlaget som foreligger. Skatteeffekten er målt, den europeiske kommisjonsulempen er det ikke, og markedsvalet står som en åpen metodisk post. Cross-sectional momentum på amerikanske enkeltaksjer er tatt inn igjen i strategiutvalget, og begrunnelsen for at testene kjøres på amerikansk marked hviler i hovedsak på datatilgang, som behandlet i kapittel 3.2.1. Strategien fungerer samtidig som kontrollgruppe for multifaktorstrategien i kapittel 2.3.6, ved at den viser hva momentum alene gir på det samme universet.

2.3.4 Faktor-ETF rotasjon for europeiske markeder

Inspirert av Asness, Moskowitz og Pedersen (2013) som dokumenterte at faktorpremier som value og momentum eksisterer på tvers av asset classes og markeder, og av nyere forskning på faktormomentum, ble en rotasjonsstrategi på europeiske faktor-ETF-er identifisert som en pragmatisk og skattemessig optimal tilnærming.

Strategien bruker UCITS-hjemmehørende faktor-ETF-er som er innenfor fritaksmetoden for et norsk investeringsselskap. Tre varianter ble formulert og testet:

- Variant 1 - Pure relative momentum: Månedlig rebalansering der man holder lik vekt i de to ETF-ene med høyest 6-måneders avkastning.

- Variant 2 - Faber-style (absolutt + relativ momentum): Som Variant 1, men kun ETF-er over sin 200-dagers SMA er kvalifisert. Hvis færre enn to kvalifiserer, holdes resten i obligasjons-ETF.
- Variant 3 - Volatilitetsvektet med momentum-tilt: Vektingen av kvalifiserte ETF-er er invers proporsjonal med deres volatilitet, justert for momentum-rangering.

2.3.5 Absolutt momentum (trendfølging)

En enklere variant av trendfølging som klassisk er beskrevet av Faber (2007) og videreutviklet av Antonacci (2014). Strategien holder en bred markedsportefølje (representert ved IMEU ETF) så lenge denne ligger over sitt langsiktige glidende gjennomsnitt, og bytter til en defensiv allokering (statsobligasjoner) når markedet bryter under. Strategien er motivert av observasjonen at de fleste store markedstap skjer i langvarige nedtrender som er identifiserbare i forkant.

Strategien tjener flere formål i denne studien: den fungerer som en enkleste-mulig benchmark for trendfølgingsprinsippet, og dens resultat kan brukes til å vurdere om kompleksiteten i faktorrotasjonsstrategiene gir noe ekstra utover en enkel switch mellom aksjer og obligasjoner.

2.3.6 Multifaktorstrategi (Quality-Value-Momentum) på enkeltaksjer

Forskningen på europeiske markeder, blant annet dokumentert av Figuerola-Ferretti, Bermejo Climent, Santos Moreno og Hevia (2021), som dekker perioden 1991-2019 på europeiske store-cap aksjer, indikerer at multifaktortilnærminger som kombinerer kvalitet, verdi og momentum på enkeltaksjer historisk har gitt sterkere risikojustert avkastning enn enkelt-faktor strategier.

Strategien opererer på følgende prinsipp: hver aksje i universet gis en standardisert score på hver av de tre faktorene, scorene kombineres til én samlet rangering, og porteføljen holder de høyest rangerte selskapene med lik vekt. Kvalitet måles ved lønnsomhet og gjeldsgrad, der bruttofortjeneste mot eiendeler følger Novy-Marx (2013), verdi ved prisingsmultipler, og momentum ved avkastning over 12 måneder eksklusiv siste måned. Standardiseringen gjøres innenfor sektor, slik at rangeringen ikke i praksis blir en sektorallokering. Porteføljen rebalanseres halvårlig.

En test av QVM stiller strengere krav til datagrunnlaget enn de øvrige strategiene i denne rapporten. Kravene er point-in-time fundamentaldata og et aksjeunivers uten survivorship bias. Disse kravene lar seg oppfylle for amerikanske aksjer innenfor prosjektets rammer, men ikke for europeiske. Sammen med kostnadsvurderingen i kapittel 3.1.3 er dette begrunnelsen for at strategien testes på amerikansk marked. Datakravene er behandlet i kapittel 3.2.1.

2.4 Begrunnelse for strategiutvalget og dets begrensninger

Strategiene som er valgt for testing er ikke et tilfeldig utvalg. De representerer en logisk progresjon: fra den opprinnelige hypotesen (RSI mean-reversion), via inkrementelle forbedringer (filtrerte varianter), til strukturelt forskjellige tilnærminger (cross-sectional momentum, faktorrotasjon, multifaktor på enkeltaksjer). Denne progresjonen reflekterer hvordan empirisk forskning typisk

utvikler seg: en hypotese testes, og dersom den ikke holder, søker man enten å forfine den eller å utforske beslektede hypoteser.

Til sammen dekker strategiene de fem hovedkategoriene innenfor systematisk handel som er praktisk gjennomførbare for en retailaktør: mean-reversion (intraday RSI), kryssektoral momentum (cross-sectional på enkeltaksjer), tidsserie momentum (faktorrotasjon), trendfølging (absolutt momentum), og multifaktor (QVM). Dette gir rapporten bredde i tilnærmingen, selv om hver enkeltkategori ikke testes uttømmende.

Det er flere strategier som av ressursmessige grunner ikke er testet fullt ut i denne rapporten:

- Statistical arbitrage og pair trading: Krever sofistikert statistisk analyse og infrastruktur for kontinuerlig overvåkning av et stort antall aksjepar.
- Carry trades i valuta: Krever toleranse for ekstrem volatilitet og er ikke umiddelbart relevant for HULC AS sin profil.
- Volatilitets-strategier: Salg av volatilitet via opsjoner og varianter krever spesialisert kompetanse og opsjons-konto.
- Sesongstrategier: Effekten er typisk så liten at de fungerer best som filtre på toppen av andre strategier.
- Mer eksotiske momentum-varianter som residual momentum, factor momentum og industry momentum.

Det er videre bevisst valgt å ikke teste alle tenkelige parameter-varianter for hver strategi. En slik tilnærming ville medføre betydelig risiko for data mining, der man ved tilstrekkelig mange forsøk vil finne noen varianter som fremstår som lønnsomme ved tilfeldighet alene (Harvey, Liu og Zhu, 2016; Bailey mfl., 2014). I stedet er hver variant motivert av spesifikke teoretiske eller praktiske begrunnelser før den testes, og parameterrobusthet undersøkes systematisk for den mest lovende strategien.

Den samlede konsekvensen av disse avgrensningene er at rapporten kan trekke konklusjoner om de testede strategiene under de testede betingelsene, men ikke om hele klassen av algoritmiske strategier i sin alminnelighet. Vi har vært innom alle de overordnede tilnærmingene, men har ikke per dags dato ressursene til å gå fullstendig i dybden på hver enkelt overordnet strategi. Basert på dette legges grunnlaget for konklusjonen som presenteres i kapittel 6, der vi vil oppsummere hva resultatene faktisk forteller oss og hva som er den rasjonelle veien videre for HULC AS gitt denne kunnskapen.

3. Metodikk

3.1 Backtestrammeverket

Det er utviklet et eget backtestrammeverk i Python for å teste de aktuelle strategiene under konsistente forhold. Rammeverket er strukturert modulært med separate komponenter for datalast, strategiimplementering, simulering av ordreutførelse, porteføljestyling og rapportgenerering. All

koden er versjonskontrollert, og hver komponent er dekket av enhetstester som kjøres i sin helhet før hver måling.

Rammeverket håndhever flere fundamentale prinsipper som er nødvendige for at backtestresultater skal være troverdige:

3.1.1 Håndhevelse av no-look-ahead

Et av de vanligste problemene i amatør-backtester er såkalt look-ahead bias, der strategien ved et uhell får tilgang til informasjon som ikke ville vært tilgjengelig på handelstidspunktet. I rammeverket er dette håndhevet teknisk. Strategien får kun se data fra perioden strengt før den aktuelle handelsdagen. Det er skrevet en spesifikk test som forsøker å bryte dette prinsippet ved hjelp av en spion-strategi, og som verifiserer at rammeverket faktisk fanger forsøket.

Prisdata er imidlertid bare én av tre kilder til look-ahead for strategier som opererer på enkeltaksjer. De to øvrige er restatements, altså at regnskapstall senere korrigeres og at den korrigerte versjonen leses tilbake i tid, og publiseringsforsinkelse, altså at et regnskapstall gjøres tilgjengelig for strategien fra periodeslutt i stedet for fra faktisk publiseringsdato. En tredje beslektet feilkilde er universkonstruksjon, der selskaper som senere ble avnotert utelates fra de periodene de faktisk var indeksmedlemmer. Ljungqvist, Malloy og Marston (2009) dokumenterer hvordan historiske databaser skrives om i ettertid uten at det er synlig for brukeren.

Rammeverket er derfor utvidet med en egen kontrakt for tilgang til fundamentaldata og universoppslag. Kontrakten krever at enhver forespørsel oppgir en beslutningsdato, at kun observasjoner som var tilgjengelige på eller før denne datoen returneres, at kun opprinnelig rapporterte verdier er synlige, og at universet bygges på historisk indeksmedlemskap. Manglende data gir eksplisitt feilmelding i stedet for tomt resultat, slik at fravær ikke kan forveksles med en gyldig observasjon. Kontrakten er leverandøruavhengig, slik at leverandørspesifikk logikk holdes i et adapterlag og ikke i strategiene.

Hva kontrakten gjør i praksis, lar seg vise på ett selskap. Lehman Brothers var medlem av S&P 500 til og med snapshotet 30. juni 2008, og begjærte seg konkurs 15. september samme år. På beslutningsdatoen 30. juni 2008 svarte kontrakten med en egenkapitalavkastning på 15,4 %, et resultat på 3 461 millioner dollar og en egenkapital på 24 832 millioner dollar. De to foregående årene ga 20,5 % og 21,2 %. Ti uker etter den siste observasjonen fantes ikke selskapet.

En kvalitetsrangering den dagen skulle plassert Lehman blant de bedre navnene i universet. Det er det riktige svaret, fordi det var informasjonen som forelå. En backtest som i stedet hadde brukt de senere korrigerte tallene, eller som hadde utelatt Lehman fra universet fordi selskapet ikke finnes i dag, ville rapportert en avkastning ingen investor kunne ha oppnådd.

Eksempelet viser publiseringsforsinkelsen like konkret. Tallene som var tilgjengelige 30. juni 2008, gjaldt regnskapsperioden som endte 29. februar 2008, og ble tilgjengelige 9. april. De var 82 dager gamle på beslutningsdatoen. Å filtrere på periodeslutt i stedet for tilgjengelighetsdato ville gitt strategien tall den ikke kunne ha sett.

3.1.2 Realistisk kostnadsmodellering

Alle simulerte handler belastes med kostnader som reflekterer faktiske vilkår hos Interactive Brokers Pro (Interactive Brokers, u.å.):

- Kommisjon: 0.005 USD per aksje, minimum 1.00 USD per ordre, maksimalt 1 % av handelsverdien.
- Bid-ask spread: 3 basispunkter per side, modellert som at kjøp skjer over og salg under midtprisen.
- Slippage: 5 basispunkter per side, i tillegg til spread. Dette er et konservativt estimat basert på faktiske handler i likvide US-aksjer og europeiske ETF-er.

Det totale kostnadsbildet per round-trip handel ligger dermed på ca. 16 basispunkter pluss flat kommisjon på små ordre. Dette er bevisst satt noe konservativt for å unngå overoptimistiske resultater. En tidligere versjon av rammeverket benyttet en mer aggressiv slippage-modell med volatilitets-koeffisient, men denne ble revidert da det viste seg at den introduserte uforholdsmessig store kostnader for ETF-strategiene. Betydningen av minimumskommisjonen ved små ordrestørrelser er behandlet i kapittel 3.1.3.

3.1.3 Skatt og strukturelle handelskostnader

Et viktig metodisk hensyn er at HULC AS som norsk aksjeselskap er underlagt selskapsbeskatning. For aksjeinvesteringer utenfor EØS gjelder en effektiv skattesats på 22 % av realisert gevinst. For aksjer og fond hjemmehørende innenfor EØS (under fritaksmetoden) er gevinstbeskatning i hovedsak frafalt for selskapsinvestorer, jf. skatteloven § 2-38.

Dette skillet har vesentlige konsekvenser for valg av strategi og marked, og var utslagsgivende for at prosjektet tidlig flyttet fra amerikanske til europeiske instrumenter. Isolert sett innebærer skatteforskjellen at en bruttoavkastning på 12 % CAGR reduseres til 9,36 % CAGR dersom hele gevinsten realiseres løpende utenfor fritaksmetoden.

Skatt er imidlertid bare én av flere strukturelle kostnader, og den er ikke nødvendigvis den største ved den kapitalstørrelsen HULC AS opererer med. To forhold gjør bildet mer sammensatt enn den opprinnelige vurderingen la til grunn.

For det første er fritaksmetoden ikke fullt ut dekkende for et europeisk aksjeunivers. Fritaksmetoden gjelder selskaper hjemmehørende innenfor EØS. Storbritannia er ikke lenger medlem av EØS, og Sveits har aldri vært det. Disse to markedene utgjør en betydelig andel av Stoxx 600. Et bredt europeisk enkeltaksjeunivers gir dermed en effektiv skattesats i størrelsesorden 7 til 8 % på realisert gevinst, ikke null. Skattefordelen mot amerikansk marked er reell, men vesentlig mindre enn 22 prosentpoeng.

For det andre er kommisjonsstrukturen forskjellig mellom markedene. Interactive Brokers opererer med et minimumsbeløp per ordre. På amerikanske aksjer er minimum 1,00 USD. På europeiske børser er minimumsbeløpet flere ganger høyere, og britiske aksjer belastes i tillegg med stamp duty på 0,5 %

ved kjøp. Minimumsbeløpet er en fast kostnad per ordre, og utgjør derfor en høyere andel jo mindre ordren er.

Konsekvensen er at rangeringen mellom markedene avhenger av ordrestørrelse. Med en allokering på 50 000 NOK til enkeltaksjestrategier fordelt på 10 til 20 posisjoner blir hver ordre i størrelsesorden 2 500 til 5 000 NOK. Ved slike ordrestørrelser utgjør minimumskommisjonen en høy andel av ordreverdien, og et europeisk univers vil bære et kostnadstrekk et amerikansk ikke har.

Den opprinnelige vurderingen la til grunn at dette kostnadstrekket overstiger skattefordelen med god margin, og at markedssvalget dermed fulgte av kostnadsbildet. Den vurderingen bærer ikke alene. Skatteeffekten er nå målt eksplisitt i porteføljelaget, og den er vesentlig større enn hele kostnadsmodellen ved de kapitalstørrelsene det er tale om. Størrelsesforholdet mellom de to avhenger dessuten av hvor mye porteføljen omsetter, og er derfor ikke en konstant som kan brukes til å rangere markeder uavhengig av strategi.

Spørsmålet er heller ikke avklart, og det av to grunner. Den europeiske kommisjonsulempen er ikke målt. Minimumsbeløpene hos Interactive Brokers og stamp duty på britiske aksjer er kjent, men det er ikke kjørt en simulering på et europeisk univers med europeiske kommisjonssatser, og det er derfor ukjent hvor stor ulempen blir uttrykt i årlig avkastning. Dernest ville en europeisk portefølje hatt en annen skatteposisjon under fritaksmetoden, med en effektiv sats i størrelsesorden 7 til 8 % i stedet for 22 %. Den målte skatteeffekten er beregnet under 22 % og lar seg ikke uten videre overføre til et europeisk univers.

Enkeltaksjestrategiene testes derfor på amerikansk marked, med 22 % skatt på realisert gevinst modellert eksplisitt i backtesten. Begrunnelsen for dette valget hviler nå i hovedsak på datatilgang, som behandlet i kapittel 3.2.1, og ikke på kostnadsbildet. Den passive delen av porteføljen holdes fortsatt i EØS-hjemmehørende UCITS-fond under fritaksmetoden. Terskelen på 300 000 NOK som tidligere er oppgitt for å revurdere markedssvalget, er ikke utledet av måling. Den bør erstattes av en beregnet terskel når den europeiske kommisjonsulempen er målt på samme måte som skatteeffekten er målt. Inntil da står markedssvalget som en åpen metodisk post.

3.2 Datakilder og datakvalitet

Forskjellige strategier krever ulike datakilder. Følgende kilder er benyttet:

- Intraday data (5-minutters bars) for US-aksjer: hentet via Interactive Brokers API. Dette er den samme datakilden som ville bli brukt i live-handel, noe som sikrer konsistens mellom backtest og potensiell produksjonsbruk.
- Daglige data for US-aksjer og europeiske ETF-er: hentet via yfinance (Aroussi, u.å.), et åpent Python-bibliotek som benytter Yahoo Finance som underliggende kilde. Auto-justering for utbytter og splits er aktivert.
- Fundamentaldata og historisk indeksmedlemskap for amerikanske enkeltaksjer: Sharadar Core US Equities Bundle, distribuert gjennom Nasdaq Data Link. Tabellen SF1 gir regnskapstall både som opprinnelig rapportert og i senere korrigert form, med

tilgjengelighetsdato per observasjon, og SP500 gir historiske inn- og utmeldinger av indeksen. Avnoterte selskaper er inkludert. Datasettet er anskaffet og i bruk, og målingene i kapittel 3.2.1 og 3.2.2 er kjørt mot det. Kravene til denne datatypen er behandlet i kapittel 3.2.1.

All data caches lokalt som Parquet-filer for å sikre reproduserbarhet og redusere belastningen på datakildene. For Interactive Brokers krever historiske intraday-forespørsler hensyn til rate limits (maksimalt 60 forespørsler per 10 minutter), og det er implementert en semafor-basert pacing-mekanisme for å overholde disse begrensningene.

Det er identifisert flere kvalitetsutfordringer ved dataene som har konsekvenser for testperioden og tolkningen av resultater:

- Momentum-ETF-en IEMO har komplett data først fra 2018; perioden 2015-2017 er tynn ved lansering.
- Visse ticker-substitusjoner var nødvendige fordi enkelte ETF-er ikke har tilstrekkelig historikk på London Stock Exchange. For eksempel ble IEMO.L erstattet med IEMO.MI (Milan-noteringen), og IMEU.L ble erstattet med IMEU.MI fordi London-noteringen kvoterer i GBp og dermed gir feilaktige prisbevegelser.
- Faktorrotasjonsstrategien er derfor testet på perioden 2018-01 til 2024-12, ikke fra ETF-enes opprinnelige lanseringsdato. Dette er gjort for å sikre at alle strategivarianter kjører på samme datagrunnlag.

3.2.1 Datakrav for enkeltaksjestrategiene

De to strategiene som opererer på enkeltaksjer stiller strengere krav til datagrunnlaget enn ETF-strategiene. Begge krever et univers med historisk indeksmedlemskap, og multifaktorstrategien krever i tillegg historiske fundamentaldata over flere år. Datakravet behandles derfor separat her, fordi det avgjør både om strategiene lar seg teste på en måte som gir tolkbare resultater, og i hvilket marked testene kan gjennomføres.

Fire krav må være oppfylt for at testene skal kunne konkluderes på. De to første gjelder begge enkeltaksjestrategiene, de to siste er særlig kritiske for multifaktorstrategien:

- Point-in-time fundamentaldata. Regnskapstall slik de var publisert på beslutningstidspunktet, ikke slik de fremstår i dag. Gratis kilder viser i to av tre tilfeller den sist korrigerte versjonen tilbake i tid, som målt nedenfor. Strategien får dermed implisitt kjennskap til korreksjoner som ikke var offentlig kjent da beslutningen ble tatt.
- Et aksjeunivers uten survivorship bias. Universet må bygges på historisk indeksmedlemskap, slik at selskaper som senere gikk konkurs, ble kjøpt opp eller ble avnotert inngår i de periodene de faktisk var medlemmer. Et univers satt sammen av dagens indeksmedlemmer består per definisjon av overlevende, og gir kunstig gode resultater uavhengig av strategiens kvalitet (Brown mfl., 1992). Dette kravet gjelder også for rene prisbaserte strategier.
- Tilstrekkelig historikk. Verdifaktoren hadde en uvanlig svak periode i utviklede markeder fra omtrent 2010 til 2020 (Fama og French, 2021). En testperiode på fem til seks år måler ett regime, og gir ikke grunnlag for å konkludere om faktoren generelt.

- Konsistent beregning av nøkkeltall. Kvalitets- og verdimålene må være beregnet likt på tvers av selskapene, og definisjonene må være de samme i backtesten som i eventuell senere drift. Avviker definisjonene, kjøres ikke den strategien som ble testet.

Kravene over er egenskaper ved datasettet, ikke ved leverandøren. Det som avgjør om en fundamental backtest kan konkluderes på, er om tallene er punkt-i-tid og om universet er fritt for survivorship bias, ikke hvilket navn som står på fakturaen. De institusjonelle leverandørene skiller seg fra det datasettet som er brukt her først og fremst på bredde i markeder og instrumenter, og på pris. Begge de to kravene er derfor testet direkte mot det datasettet som faktisk er brukt, framfor å forutsettes oppfylt fordi kilden er anerkjent. Målingene nedenfor er den verifikasjonen.

Påstanden om at gratis kilder ikke holder, er testet direkte mot det kommersielle datasettet. Utvalget er 120 tilfeldig trukne S&P 500-medlemmer per medlemsdato, til sammen 311 unike symboler. Andelen av datidens indeksmedlemmer som yfinance i dag ikke har noe som helst på, faller monotont med alderen på medlemskapet.

Medlemsdato	Utvalg	Mangler	Andel
30.06.2005	120	54	45,0 %
30.06.2015	120	24	20,0 %
30.06.2022	120	10	8,3 %

Mellom 92 og 96 % av de manglende selskapene er avnoterte. Dette er survivorship i sin reneste form. Kilden har i praksis bare selskapene som fortsatt finnes. En backtest fra 2005 bygget på yfinance ville manglet nesten halve universet, og de manglende er systematisk de som gikk dårligst. Dybden er dessuten fire år og ikke fem. yfinance returnerer fem årskolonner, men den eldste er tom, og et forsøk på regnskapsår 2021 ga null sammenlignbare observasjoner av 595 mulige. Kilden kan derfor ikke brukes til fundamental backtesting utover fire år, uavhengig av hvordan korreksjoner håndteres.

Blant selskapene som finnes, avviker yfinance-verdien fra det opprinnelig rapporterte tallet med mer enn 0,5 % i 57 av 448 tilfeller for regnskapsåret 2022, altså 12,7 %, og i 64 av 537 for 2023, altså 11,9 %. Blant dem som avviker er medianavviket 3,06 %, p95 er 23,83 % og maksimum 88,3 %.

Avviksraten varierer sterkt mellom felt, og skillet er avgjørende for tolkningen. For samlet gjeld avviker 74 av 197 observasjoner, altså 37,6 %, men bare to av dem treffer den korrigerte verdien. Det er en definisjonsforskjell og ikke en tidsforskjell. Kildene behandler blant annet leieforpliktelser ulikt. For omsetning avviker 38 av 197, altså 19,3 %, og 16 av dem treffer den korrigerte verdien. For eiendeler, egenkapital og resultat er definisjonene sammenlignbare, og avviksratene er henholdsvis 0,5 %, 1,0 % og 3,0 %.

For de tre feltene der definisjonene er sammenlignbare, avviker 9 av 591 observasjoner, altså 1,5 %, og samtlige ni treffer den korrigerte verdien. Snudd om. Blant de 42 observasjonene der det kommersielle datasettet selv har registrert en korreksjon, viser yfinance den korrigerte verdien i 27 tilfeller, altså 64,3 %, den opprinnelig rapporterte i 11 tilfeller, altså 26,2 %, og noe annet i 4 tilfeller. Treffene er eksakte og ikke omtrentlige. Kilden leverer nøyaktig det tallet selskapet oppga senere.

En backtest på yfinance rammes dermed av to feil samtidig, og begge trekker samme vei. Universet mangler taperne, og der en korreksjon finnes, viser kilden den i to av tre tilfeller. Den første feilen gir for høy avkastning fordi konkursene er borte. Den andre gir for gode faktorverdier fordi regnskapstallene er de reviderte. Ingen av dem gir seg til kjenne i resultatet. Begge ser ut som en bedre strategi.

Hvor mye korreksjoner betyr, er målt på datasettets egen as-reported-dimensjon, avgrenset til selskaper som har vært S&P 500-medlemmer i perioden 1998 til 2026. Av 91 740 regnskapsperioder har 1 321 mer enn én innsending, altså 1,44 %. Målt per selskap er 724 av 1 159 berørt minst én gang, altså 62,5 %. Medianen fra opprinnelig innsending til korreksjon er 23 dager, p95 er 106 dager og maksimum 452 dager.

Kontrasten mellom de to andelene er poenget. Korreksjoner er sjeldne per regnskapsperiode og nesten universelle over et selskaps levetid. En backtest som leser korrigerede tall, får derfor ikke en liten og jevnt fordelt feil. Den får riktige tall for de fleste periodene, og systematisk for gunstige tall for nettopp de periodene der noe gikk galt.

Størrelsen på korreksjonene peker samme vei. Av de seks tilfellene testsuiten bruker, snur fire enten fortegn eller endrer resultatet med mer enn 90 %. Tyco Internationals resultat for kvartalet som endte 31. mars 2002, ble innsendt 15. mai som pluss 1 400 millioner dollar og korrigeret 12. juni til minus 3 113 millioner, altså 28 dager senere. En kvalitetsfaktor beregnet på korrigerede tall ville rangert Tyco korrekt som et av universets svakeste selskaper våren 2002. Den rangeringen fantes ikke da beslutningen skulle tas, og en backtest som bruker den, måler evnen til å se inn i framtiden.

Kravene lar seg derimot oppfylle for amerikanske aksjer. Kommersiell point-in-time datasett med avnoterte selskaper inkludert, historikk tilbake til 1990-tallet og egen tabell over historisk indeksmedlemskap er tilgjengelige til en månedlig kostnad som ligger godt innenfor prosjektets rammer. Tilsvarende dekning for europeiske aksjer krever institusjonelle databaser som ikke er tilgjengelige her. Dette er, sammen med kostnadsvurderingen i kapittel 3.1.3, begrunnelsen for at enkeltaksjestrategiene testes på amerikansk marked.

Det bør presiseres at fundamentaldata levert gjennom meglerens markedsdataabonnement ikke dekker dette behovet. Slike tjenester gir gjeldende nøkkeltall for noterte selskaper, ikke historiske observasjoner slik de forelå på beslutningstidspunktet, og de inkluderer ikke selskaper som senere er avnotert. De er anvendelige for løpende drift, men ikke som grunnlag for en backtest.

Datatilgang alene sikrer ikke at testen blir korrekt. Slike datasett leverer regnskapstall i flere versjoner, både som opprinnelig rapportert og i senere korrigeret form, og hver observasjon er merket med datoen den ble tilgjengelig. Point-in-time-korrekthet oppnås først når spørringen både velger den opprinnelige rapporterte versjonen og filtrerer bort observasjoner som ikke forelå på beslutningstidspunktet. Feil her gir ingen synlig feilmelding, kun et for godt resultat. Rammeverkets no-look-ahead-test er derfor utvidet til å dekke fundamentaltilgang.

Selv med korrekte data avgjør ikke avkastningstallene alene om strategien har verdi. Positiv risikojustert avkastning kan skyldes eksponering mot kjente faktorer snarere enn strategiens sammensetning. Resultatene kontrolleres derfor mot en faktorregresjon, som beskrevet i kapittel 3.3.

3.2.2 Survivorship gjennom datadekning

Kontrakten stenger survivorship i universkonstruksjonen. Medlemslistene er historiske, og selskaper som gikk konkurs eller ble kjøpt opp, er med i de periodene de var medlemmer. Multifaktorstrategien ekskluderer imidlertid i tillegg selskaper der fundamentaldata mangler. Henger manglende dekning sammen med at selskapet var i ferd med å gå under, gjenoppstår skjevheten gjennom en annen mekanisme enn den kontrakten dekker. Det er målt.

Over 53 rebalanseringer, med et vindu på 12 måneder, ble 211 navn ekskludert fra strategien på grunn av manglende felt. Av disse forsvant 24 fra indeksen innen 12 måneder, altså 11,4 %. I universet samlet forsvant 1 102 av 26 497 navn, altså 4,2 %. Relativ risiko er 2,73. En tosidig z-test gir $z = 5,20$ og $p = 2,1 \times 10^{-7}$. Forskjellen er ikke forenlig med tilfeldighet. De to siste av de 55 rebalanseringsdatoene er utelatt, fordi de mangler et fullt tolv månedersvindu.

Funnet er mindre alvorlig enn tallet alene tilsier, fordi de 24 eksklusjonene fordeler seg på 15 unike selskaper og på tre ulike mekanismer.

Mekanisme	Antall	Selskaper
Utenlandsk domisil	10	UL, SHEL, INCLF, PDG, B
Restrukturering	8	BXLT, CPGX, TFCFA, H1, VNT
Reell nød	6	SBNY, FRCB, FMCC, FLTWQ, LDWIF

Utenlandsk domisil er den største gruppen og den minst bekymringsfulle. Samtlige ble fjernet 22. juli 2002, da S&P tok utenlandske selskaper ut av indeksen. De manglet US-GAAP-tall fordi de var utenlandske, og ble fjernet fordi de var utenlandske. Korrelasjonen er ekte, men den handler om regnskapsstandard og ikke om selskapets tilstand. Restrukturering dekker oppkjøp, fisjoner og fusjoner der regnskapsrapporteringen ble brutt av selve transaksjonen, og retningen der er uklar, siden oppkjøp skjer både til premie og i nødssalg. Reell nød er den gruppen som utgjør en skjevhet. Selskaper slutter å levere rapporterbare tall når de er i ferd med å gå under, og faller derfor ut av strategien rett før de kolliderer.

Retningen på skjevheten er kjent. Den favoriserer multifaktorstrategien relativt til den rene prisbaserte strategien, altså den strategien hypotesen handler om. Multifaktorstrategien får ikke lov til å eie de dårligste utfallene i nød-gruppen, fordi de er usynlige på beslutningsdatoen. Den prisbaserte strategien rangerer kun på pris og eier dem. Både FRCB og SBNY inngår i det prisbaserte universet ved 31. desember 2022, med 12-1-momentum på henholdsvis minus 37,9 % og minus 55,9 %. Sammenligningen mellom de to strategiene er dermed ikke helt nøytral, og skjevheten går i favør av den strategien som testes.

Ingen korreksjon er forsøkt. Å kompensere ville krevd antakelser om hva de manglende tallene ville vært, altså å tilskrive selskaper som sluttet å rapportere en fiktiv lønnsomhet og gjeldsgrad. Det ville erstattet en kjent og målt skjevhet med en ukjent modellfeil.

Størrelsesordenen bør stå ved siden av signifikansen. 24 selskapseksklusjoner over 53 rebalanseringer med omtrent 495 kvalifiserte navn hver utgjør i underkant av 0,05 % av alle posisjonsmuligheter. I en portefølje på 15 navn er sannsynligheten for at ett av dem faktisk ville blitt valgt, liten, og de ville uansett rangert lavt på kvalitetsfaktoren. Skjevheten er statistisk sikker, men effekten på avkastningstallene er sannsynligvis marginal. Begge deler er en del av funnet.

3.2.3 Konstruksjon av multifaktorscoren

Multifaktorstrategien er den mest datakrevende av testene, og konstruksjonen av selve scoren har flere valg som hver for seg kan endre hvilke selskaper som velges. De er samlet her, fordi et resultat ikke lar seg tolke uten dem.

Scoren bygges av åtte mål fordelt på tre blokker. Et selskap må ha ferske verdier i tolv fundamentalfelter for å i det hele tatt bli rangert, og kvalifiseringen skjer i datalaget før strategien ser universet. Et selskap som faller ut, havner i en eksklusjonsliste med begrunnelse, ikke i stillhet.

Mål	Blokk	Retning
Egenkapitalavkastning, TTM	Kvalitet	Høyere er bedre
Bruttofortjeneste / eiendeler	Kvalitet	Høyere er bedre
Gjeld / egenkapital	Kvalitet	Lavere er bedre
Nettoresultat / markedsverdi	Verdi	Høyere er bedre
Bokført egenkapital / markedsverdi	Verdi	Høyere er bedre
Driftsresultat / enterprise value	Verdi	Høyere er bedre
Fri kontantstrøm / markedsverdi	Verdi	Høyere er bedre
Avkastning 12 mnd eks. siste mnd	Momentum	Høyere er bedre

Alle mål orienteres til «høyere er bedre» ett sted i koden. Spredt sorteringsretning er den mest sannsynlige kilden til en fortegnseil som ingen test fanger, fordi et snudd fortegn gir et fullt gyldig resultat.

Verdimålene er skrevet som yield og ikke som multippel. Det er ikke kosmetikk. En multippel med negativ nevner snur fortegn. Et selskap som taper penger får negativ pris/fortjeneste, og i en sortering på lavest først havner det øverst, som universets billigste selskap. Yield-formen gir negativt resultat en negativ verdi som sorterer nederst, der den hører hjemme. Feilen er stille i multippelformen, og den rammer nettopp de selskapene et verdifilter skal luke bort. Det samme gjelder to steder til. Gjeldsgrad beregnet på negativ egenkapital er ikke lav gjeld, den er meningsløs, og driftsresultat delt på enterprise value er det når selskapet har mer kontanter enn markedsverdi. Begge settes til manglende framfor å bli lest som et kvalitetstegn.

Målene omgjøres til rangpersentiler og ikke til z-scorer. Fundamentale forholdstall har tunge haler. Én egenkapitalavkastning på 400 %, fra et selskap med nesten ingen bokført egenkapital, flytter både gjennomsnitt og standardavvik nok til å endre hele blokkens scorer. En rangering bryr seg om rekkefølgen og ikke om avstanden, og er upåvirket. Z-score finnes som alternativ med symmetrisk klipping av 1 % i hver hale, slik at valget kan testes framfor å bli postulert. Rangbasert standardisering innenfor sektor følger framgangsmåten hos Asness, Frazzini og Pedersen (2019).

Rangeringen skjer innenfor sektorgruppe. Uten det ville verdiblokka i praksis blitt et sektorveddemål. Banker har strukturelt lav pris mot bokført egenkapital og programvare strukturelt høy, og en ren tverrsnittsrangering ville fylt porteføljen med finans i enhver periode.

Valget av klassifisering er begrunnet i en måling og ikke i en preferanse. Ingen av de tilgjengelige klassifiseringene har historikk, så stabilitet over tid kan ikke måles direkte. Det som kan måles, er hvor mye selvstendig skjønn hver av dem inneholder, målt mot bransjekoden selskapet selv oppgir til amerikanske myndigheter.

Klassifisering	Kategorier	Kode gir flere verdier
SIC-sektor	9	0,0 %
Fama-French 48	48	1,1 %
Morningstar-sektor	11	58,4 %

De to første er i praksis deterministiske funksjoner av en innrapportert kode og arver dens stabilitet. Morningstars sektorinndeling er det ikke: 58 % av kodene fordeler seg på flere sektorer, noe som berører 86 % av selskapene. Den bærer altså redaksjonelt skjønn som kan revideres uten spor, og en revisjon ville endret en historisk backtest.

Valget er Fama-French tolv industrier (Fama og French, 1997), utledet fra bransjekoden etter Frenchs publiserte definisjon (French, u.å.). Den arver kodens stabilitet, den er inndelingen litteraturen bruker, og den kan siteres. Inndelingen med 48 industrier ville vært for fin: 48 grupper over omtrent 500 indeksmedlemmer gir ti navn per gruppe, og halen er verre. Blant selskaper som noen gang har vært indeksmedlem har én gruppe ett navn og to grupper har to. En rangering innenfor en gruppe på ett navn måler ingenting. Med tolv industrier blir gruppene mellom 7 og 100 navn med median 36, og ingen gruppe faller under fem navn ved noen rebalansering. Mekanismen som slår sammen for små grupper er likevel beholdt, fordi et univers en gang kan bli smalere enn dagens.

Klassifiseringen har en kjent skjevhet som ikke lar seg fikse med disse dataene. Tabellen den hentes fra er et nåsnapshot uten historikk, med én rad per selskap og ingen opplysning om når klassifiseringen ble satt. Et selskap som byttet bransje i 2007 bærer 2026-verdien også i 1999-rebalanseringen. Retningen på skjevheten er verdt å merke seg. Sektornøytralisering med etterpåklok klassifisering plasserer et selskap i den bransjen det endte i, og sammenligner det med feil konkurrenter i de tidlige periodene. Leverandøren tilbyr ingen historisk sektorserie, og skjevheten er derfor dokumentert og ikke rettet.

Selskaper med flere aksjeklasser krever en egen bro. Indeksen inneholder begge klassene til enkelte selskaper, og to ting går galt hvis de behandles som to selskaper. Regnskapet leveres bare under primærklassen, så sekundærklassen ville falt ut som manglende data selv om tallene finnes. Og i en likevektet portefølje på 15 navn ville selskapet fått 2/15 i stedet for 1/15, altså 6,7 % av porteføljen tildelt på grunn av en kapitalstruktur og ikke et signal. Broen finner det symbolet som faktisk bærer fundamentaldataene, og dedupliserer på det.

Broen løser ikke tilfellet der bæreren av regnskapstallene selv ikke er indeksmedlem. Da faller selskapet ut helt. Det telles og logges per rebalansering, og tellingen ga treff med en gang. Twenty-First Century Fox fører regnskapet under et symbol som ikke var medlem i perioden, og selskapet falt derfor ut av sju rebalanseringer fra desember 2015 til desember 2018. Ett navn av omtrent 495 påvirker ikke resultatet nevneverdig. Poenget er at antakelsen om at broen alltid holder, ville vært feil, og at det ble oppdaget fordi det ble talt.

Til slutt kreves det at alle tre blokkene har minst ett beregnbart mål. Uten det kravet ville et selskap uten verdiblokk blitt scoret på to av tre blokker og sammenlignet med selskaper scoret på tre, altså en sammenligning der vektene betyr forskjellige ting fra selskap til selskap. Alternativet, å gi manglende blokker medianscore, er verre. Medianen ligger midt på treet, og et selskap ville da vunnet på å mangle data hver gang blokka ville trukket det ned.

Blokkvektene er like. Det finnes ingen måling i dette prosjektet som begrunner en skjev vekt, og en optimalisert vekt ville vært tilpasset samme historikk som senere skal evaluere den. Lik vekt i koden viser seg imidlertid ikke å bli lik vekt i porteføljen, av en grunn som først ble synlig da scoren ble målt. Det er behandlet i kapittel 4.7.

3.3 Statistiske mål

Resultatene evalueres ved hjelp av et sett standardiserte mål som er etablerte i kvantitativ finansiell forskning. Hvert mål adresserer en spesifikk dimensjon av strategiens ytelse:

- Total return og CAGR (Compound Annual Growth Rate): Måler absolutt avkastning over perioden og annualisert vekstrate. Sentralt for å vurdere strategiens evne til kapitalvekst.
- Sharpe ratio (Sharpe, 1994): Måler avkastning per enhet total volatilitet. Lar oss sammenligne strategier med ulike volatilitetsnivåer.
- Sortino ratio (Sortino og Price, 1994): Som Sharpe, men kun nedside-volatilitet straffes. Reflekterer at oppside-svingninger ikke er uønsket.
- Maksimal drawdown: Største kumulative tap fra topp til bunn. Reflekterer den verste opplevde nedturen og er kritisk for å vurdere strategiens psykologiske gjennomførbarhet.
- Calmar ratio: CAGR delt på maksimal drawdown. Et mål på avkastning relativt til den verste nedturen.
- Profit factor: Total gevinst delt på totalt tap. Mål på hvor mye man tjener på vinnere relativt til hva man taper på tapere.

- Win-rate: Andel av handler som er lønnsomme. Mindre informativt isolert, men relevant kombinert med profit factor.

Resultatene sammenlignes mot to referansestørrelser: passiv kjøp-og-hold av et bredt indeks-ETF i samme marked, og en klassisk 60/40-portefølje (60 % aksjer, 40 % obligasjoner). Disse referansene er valgt fordi de representerer realistiske alternativer for kapitalen i HULC AS, og fordi det er disse en aktiv strategi må slå for å rettferdiggjøre kompleksiteten.

For strategier som opererer på enkeltaksjer suppleres disse målene med en faktorregresjon. Strategiens meravkastning regresseres mot etablerte faktorserier for marked, størrelse, verdi, lønnsomhet, investering og momentum (Fama og French, 2015; Carhart, 1997). Formålet er å skille mellom avkastning som skyldes eksponering mot kjente faktorer, og som kan oppnås billigere gjennom indeksfond, og avkastning som ikke lar seg forklare av disse. Et positivt konstantledd som ikke er statistisk signifikant tolkes som fravær av dokumentert edge.

Turnover oppgis tosidig i denne rapporten, altså som summen av kjøp og salg målt mot porteføljeverdien. En tosidig turnover på 200 % betyr at hele porteføljen er byttet ut én gang. Konvensjonen står her fordi den ikke er entydig i litteraturen, og fordi de to motorene i rammeverket regner den ulikt, porteføljemotoren tosidig og ETF-motoren ensidig. Turnover-tall fra ETF-strategiene i kapittel 4.3 til 4.5 må derfor dobles før de sammenlignes med tallene for enkeltaksjestrategiene.

3.4 Validering og kvalitetssikring

For å sikre at resultatene er troverdige er flere kontrollmekanismer implementert:

- Enhetstesting:

Enhetstestene dekker kritiske komponenter som RSI-beregning, kostnadsmodellering, no-look-ahead-håndhevelse for både pris- og fundamentaldata, ordresimulering og rapportgenerering.

- Sanity checks underveis: Tidlig i prosjektet ble en bug i slippage-modelleringen identifisert da en backtest viste en 0 % win-rate over hundrevis av handler, et statistisk usannsynlig resultat. Dette ledet til revisjon av modellen og illustrerer hvordan kritisk gjennomgang av uventede resultater er en del av metoden.
- Mutasjonstesting av kontrollmekanismene: En testsuite kan være grønn på en regel den aldri utøver. For å avdekke dette er kjente feil innført i koden én om gangen, og det er verifisert at riktig test faktisk ble rød i hvert tilfelle. Feilene ble deretter fjernet og ren tilstand verifisert. Øvelsen er kjørt i tre runder. Først mot den syntetiske fixture-suiten, med seks innførte look-ahead-feil. Deretter mot suiten som kjører på det kommersielle datasettet og på porteføljelaget, med tretten innførte feil som dekker look-ahead, kolonneforveksling, oppgjørsrekkefølge, ordrefylling, kommisjonsgulv, skattebehandling, verdsettelse av beholdningen og avvikling av avnoterte posisjoner. Til slutt mot multifaktorscoreren, med elleve innførte feil i fortegn, aggregering, filtrering og gruppering. Ingen av de siste elleve kaster feilmelding; alle gir et gyldig, men galt, resultat.
- Konservante antakelser: Der det er rom for tolkning er den konservative tilnærmingen valgt. Eksempelvis antas det at når både stop loss og take profit ligger innenfor samme dagsbars

range, antas det at stop loss ble truffet først. Dette er pessimistisk, men ærlig på daglig data der intradag-rekkefølge ikke kan avgjøres.

- Eksplisitt dokumentasjon av begrensninger: Hvert resultat er dokumentert med konkrete forbehold knyttet til perioden testet, antall handler, og kjente metodiske svakheter.

Mutasjonstesting avdekket et mønster som er viktigere enn de enkeltfeilene den fant. Flere ganger har en test vært grønn på en regel den ikke utøvde, og fellesnevneren er den samme hver gang. Assertionen var riktig, og den var stilt mot riktig objekt. Det som sviktet, var at testscenariet ikke plasserte systemet i en tilstand der feilen kunne oppstå. Fire tilfeller viser formene svikten tar.

Feil dato. Den opprinnelig planlagte restatement-testen besto selv når versjonsfilteret ble fjernet helt. Den spurte på en dato der korreksjonen ennå ikke var publisert, og en versjonsblind implementasjon returnerer da den opprinnelige verdien uansett. Regelen ville vært utestet, med en grønn suite som sa det motsatte. Testen ble skrevet om til å spørre på en dato etter at korreksjonen ble tilgjengelig.

Feil oppløsning. Mutasjonen som regnet momentum på ujustert kurs i stedet for utbyttejustert, ga null røde tester. Testen fantes, og den sammenlignet momentum mot den utbyttejusterte serien. Men toleransen var 0,03, mens den faktiske forskjellen mellom de to seriene for testselskapet var 0,026. Testen målte riktig størrelse med en oppløsning som ikke skilte riktig fra galt. Den ble skrevet om til fire selskaper med høy direkteavkastning, og med den avgjørende påstanden snudd. Momentumtallet må ligge langt fra den utbyttefrie serien, ikke bare nær den justerte.

Feil lag. Testen på sektornøytralisering viste at scorefunksjonen nøytraliserer riktig når den får sektorgrupper. Den sa ingenting om hvorvidt strategien faktisk sender grupper. Mutasjonen lot funksjonen være i fred og fikk strategien til å slutte å kalle den riktig, og suiten var fortsatt grønn. En test på en ren funksjon tester funksjonen, ikke at den brukes.

Dokumentasjon mot kode. Den fjerde er av en annen art, og den er den alvorligste. Rutinen for årsoppgjør av skatt hentet årets realiserte gevinst ut av bøtta, og solgte deretter unna nok til å dekke skattekravet. Gevinsten fra de salgene ble skrevet tilbake til samme år, en bøtte som allerede var tømt og aldri leses igjen. I testscenariet gikk over 200 000 dollar realisert gevinst ubeskattet. Modulens egen dokumentasjon beskrev den riktige oppførselen presist. Koden gjorde noe annet. Ingen sluttall så feil ut. Porteføljen vokste, skatten ble betalt, og kontantbeholdningen stemte.

Konsekvensen for arbeidsmåten er den samme i alle fire tilfellene. Ingen av dem ville blitt funnet ved kodegjennomgang. De tre første er varianter av en for slapp test, der koden var riktig og testen grønn. Det fjerde er verre, fordi både dokumentasjonen og intensjonen er riktige, og bare koden er feil, på en måte som gjør resultatet bedre. En gjennomgang av dokumentasjon mot intensjon avdekker den ikke. Alle fire ble funnet ved å innføre feilen med vilje og se om testen merket det.

Runden mot multifaktorscoreren viste det samme i skala. Fire av elleve mutasjoner gikk gjennom uten en eneste rød test i første forsøk, og fire nye tester måtte skrives mot nettopp de fire feilene. De har lite til felles i kode og alt til felles i form. Ingen av dem kaster feilmelding, ingen gir urimelige tall, og alle fire ville gitt et komplett backtestresultat som så helt normalt ut. Én av dem er verdt å nevne særskilt, fordi den er den vanligste måten en faktormodell mister integriteten sin på, nemlig å fylle

manglende data med medianen. Et selskap vurdert på to blokker sammenlignes da med selskaper vurdert på tre, og fordi medianen ligger midt på treet, vinner selskapet på å mangle data hver gang blokkene ville trukket det ned.

Et eget funn gjelder testen utforming snarere enn deres innhold. Mutasjonen som beregnet markedsverdi på ujustert kurs, gjorde bare én av de to markedsverdi-testene rød. Den brede testen kontrollerer 400 selskaper mot datasettets egen markedsverdi-kolonne med en terskel på 85 % treff, og den besto, fordi store selskaper sjelden reverssplitter. Bare den smale testen ble rød, den som kontrollerer ett selskap valgt nettopp fordi det er ekstremt. Selskapet reverssplittet med faktor 240, og med feil kolonne blir markedsverdien 1,7 millioner i stedet for 415 millioner. Halen er ikke tilfeldig. Det er nettopp de små og kriserammede selskapene som reverssplitter, og feilen ville konsekvent fått dem til å se billige ut. En verdifaktor ville plukket dem systematisk. En aggregert kontroll med terskel kan derfor ikke erstatte et konkret tilfelle valgt fordi det er ekstremt.

Øvelsen viste også at antall røde tester er et dårlig mål på hvor alvorlig en feil er. I fixture-suiten ga fjerning av versjonsfilteret én rød test, mens filtrering på periodeslutt i stedet for tilgjengelighetsdato ga ni. De to feilene er like ødeleggende for et backtestresultat. I suiteen mot det kommersielle datasettet gir de samme to feilene seks røde tester hver, altså et helt annet forholdstall for samme par av feil. Antallet sier noe om hvor mange tester som tilfeldigvis berører en kodesti, ikke om hva som står på spill. Den som prioriterer etter dekningsstall, vil feilprioritere.

Mutasjonene er valgt av den som skrev koden, og dekker derfor de feilene han kom på. Feil i selve feltoversettelsen, at et felt viser seg å måle noe annet enn antatt, fanges ikke av noen av dem, og heller ikke av suiteene for øvrig. Kontrakten garanterer at tallet var kjent på beslutningsdatoen, ikke at det er riktig.

En siste feil ble ikke funnet av noen test, men av at to rapporter ble lest mot hverandre. Porteføljemotoren traff rebalanseringsdatoen med eksakt likhet mot handelskalenderen. Falt datoene på en helg eller en helligdag, fantes den ikke i kalenderen, og rebalanseringen ble hoppet over uten logglinje. Det gjaldt 16 av 55 datoer over perioden 1999 til 2026, altså 29 %, og de berørte periodene fikk dermed tolv måneders intervall i stedet for seks. Ingen sluttall så feil ut. Feilen ble oppdaget ved at antall rader i en resultatfil ikke stemte med antall rebalanseringsdatoer oppgitt i teksten.

Rettelsen ruller en rebalansering fram til første påfølgende handelsdag. Framover og ikke bakover, fordi rangeringen er beregnet per snapshotdato. Å utføre den på siste handelsdag før datoene ville vært å handle på informasjon som ikke fantes ennå. Konsekvensen er at åtte av de seksten datoene, alle årsskifter, nå utføres i januar, slik at gevinsten realiseres i neste skatteår. Det er riktig, siden handelen faktisk skjer da. Begge enkeltaksjetestene er kjørt på nytt etter rettelsen, og resultatene i kapittel 4.6 og 4.7 er fra den rettede motoren.

To ting ved denne feilen er verdt mer enn feilen selv. Den hadde retning. De utelatte rebalanseringene gjorde momentumstrategien 1,16 prosentpoeng bedre enn den er, mens de knapt rørte multifaktorstrategien. En tilfeldig utelatelse skulle ikke hatt en systematisk retning, og årsaken er ikke målt. Og den snudde en konklusjon. Rangeringsbåndets effekt i multifaktorstrategien ble målt til

minus 0,12 prosentpoeng med 39 rebalanseringer og pluss 0,45 med alle 55. Et funn om en driftmekanisme er dermed også et funn om hvor ofte mekanismen får virke. Rettelsen er nå dekket av en egen test, som bruker en rebalanseringsdato lagt på en lørdag med vilje.

3.5 Metodikkens begrensninger

Det er nødvendig å være eksplisitt om hva backtestresultater kan og ikke kan fortelle oss. Hovedbegrensningene er:

For det første varierer testperiodene sterkt mellom strategiene. Intraday- og ETF-strategiene dekker mellom seks måneder og syv år, mens enkeltaksjestrategiene kjøres over 27 år, fra juni 1999 til juni 2026. I kvantitativ finans betraktes typisk 15 til 20 år som ønskelig for solid validering. Det kravet er innfridd for enkeltaksjestrategiene og ikke for de øvrige. Forbeholdet gjelder derfor ikke rapporten som helhet, men de testene der perioden faktisk er kort.

For det andre er backtestresultater per definisjon en simulering basert på historisk data. De reflekterer ikke psykologiske og operasjonelle faktorer som påvirker faktisk live-handel: latency, partial fills, market impact, og ikke minst handlernes egen disiplin under press. Litteraturen på backtest-overtilpasning anbefaler av denne grunn å nedjustere en observert Sharpe-ratio før den tolkes som en forventning om framtidig resultat (Harvey og Liu, 2015; Bailey mfl., 2014). Rapporten oppgir ikke et tallfestet påslag, siden størrelsen avhenger av hvor mange konfigurasjoner som er prøvd.

For det tredje må strategier som ikke har edge i den testede perioden ikke automatisk forkastes for alle tider. Markedsregimer endrer seg, og strategier kan ha edge i fremtiden selv om de ikke har det nå, og omvendt.

For det fjerde har vi som lite norsk investeringsselskap naturlige ressursmessige begrensninger. Tilgangen til historiske fundamentaldata er løst for amerikansk marked gjennom et kommersielt datasett, men tilsvarende dekning for europeiske aksjer ligger utenfor prosjektets rammer. Vi har heller ikke kostbare backtesting-plattformer eller dedikert utviklingstid på fulltid. Det vi har bygget representerer hva som er praktisk gjennomførbart innenfor disse rammene, og det reflekteres i bredden over dybden i vårt strategiutvalg.

Til slutt er det slik at om en spesifikk strategiimplementering ikke har edge, er det ikke det samme som at hele tilnærmingen er feilaktig. Det kan finnes varianter, parametre, markeder eller tidsperioder der strategien gir andre resultater. Det vi kan si er at vi har testet konkrete spesifikasjoner av strategiene under realistiske forhold, og resultatene gjelder for disse spesifikasjonene.

Det er også en grense for hva kontrollmekanismene beskrevet i kapittel 3.4 kan si noe om. Fixture-suiten kjøres mot syntetiske data og beviser at mekanismen virker, ikke at et ekte datasett har de egenskapene mekanismen forutsetter. Det finnes nå i tillegg en egen suite som kjører mot det kommersielle datasettet, og den viser at datasettet faktisk leverer opprinnelig rapporterte versjoner og reelle tilgjengelighetsdatoer. Det som fortsatt står ubesvart, er ikke tidspunktet, men verdiene. Kontrakten garanterer at et tall var kjent på beslutningsdatoen, ikke at det er riktig. Feil fortegn, feil

valuta, feil skalering eller et nøkkeltall beregnet på en annen definisjon enn antatt passerer uhindret. Kvalitetskontroll av selve verdiene er en separat oppgave som ikke er påbegynt, og utgjør en selvstendig risiko.

Med disse forbeholdene presentert, og med en ærlig anerkjennelse av både hva metoden kan og ikke kan fortelle oss, går vi nå over til den eksperimentelle delen av rapporten. Resultatene som følger må tolkes innenfor disse rammene, og konklusjonen i kapittel 6 vil oppsummere hva vi faktisk kan si og hvilken vei videre dette legger opp til for HULC AS.

To begrensninger gjelder multifaktorstrategien særskilt. Sektorklassifiseringen er et nåsnaupshot uten historikk, som beskrevet i kapittel 3.2.3, og plasserer selskaper i den bransjen de endte i. Aldersgrensen på 180 dager for fundamentaldata er heller ikke utledet av data. Den er satt fordi den dekker to kvartaler med normal publiseringsforsinkelse, og den bør revurderes mot den faktiske rapporteringsfrekvensen i datasettet. Begge er valg som står skrevet, slik at de kan overprøves.

4. Resultater

I dette kapitlet presenteres resultatene fra de syv ulike strategitestene som er gjennomført i løpet av prosjektperioden. Strategiene presenteres i den kronologiske rekkefølgen de ble testet, og hver seksjon inneholder strategispesifikke resultater, sammenligning mot relevante benchmarks, og en kort tolkning av funnene.

4.1 Intraday RSI mean-reversion (Baseline)

Baseline-strategien ble testet på 33 US large-cap aksjer over perioden januar 2024 til juni 2024. Strategien åpnet posisjoner på aksjer som hadde RSI under 30 ved børsåpning, med faste 3 % take-profit og 1.5 % stop-loss bracket-ordrer, og time-stop etter 90 minutter.

Resultatene var entydig negative, med total avkastning -13.13 %, win-rate 33 %, og 85 % av posisjonene ble lukket gjennom time-stop. Det siste tallet er særlig avslørende. Det indikerer at den forventede mean-reversion-bevegelsen i hovedsak ikke materialiserte seg innenfor strategiens 90-minutters tidshorisont.

4.2 Filtrede intraday-varianter (V2)

Den kombinerte v2-strategien (alle fem filtre aktive) endte på -1.30 %, en betydelig forbedring fra baseline men fortsatt negativ. Den isolerte filter-testen viste at gapstørrelse-fileret ga den mest lovende isolerte forbedringen (-0.41 %), mens ATR-basert SL/TP og risikobasert posisjonsstørrelse var aktivt skadelig. Hypotese 1 falsifiseres på et solid empirisk grunnlag for de testede betingelsene.

4.3 Faktor-ETF rotasjon for europeiske markeder

Tre varianter ble testet på perioden 2018-2024:

- V1 (Pure Relative Momentum, default 6mnd/K=2): +41.4 %, CAGR 5.07 %, Sharpe 0.45, max DD -20.7 %.
- V2 (Faber-style, SMA200 på hver faktor-ETF): +3.1 %, Sharpe 0.10, max DD -30.7 %. SMA-filteret aktivt skadelig.
- V3 (Vol-vektet med momentum-tilt): +16.4 %, Sharpe 0.24, max DD -29.6 %.

Til sammenligning ga buy-and-hold IMEU +53.7 % (Sharpe 0.45, DD -35.2 %) og 60/40 ga +29.8 % (Sharpe 0.42, DD -24.1 %). Ingen av variantene slo buy-and-hold på avkastning, og kun V1 matchet på Sharpe.

4.4 V1 parameterrobusthet

For å undersøke om V1-resultatet er robust til parameter-valg, ble V1 kjørt med alle 12 kombinasjoner av momentum-lookback (3, 6, 9, 12 mnd) og top-K (1, 2, 3). Forhåndsdefinert robusthets-definisjon krevde minst 75 % positive Sharpe-verdier, lav spredning, og ingen dramatisk drawdown.

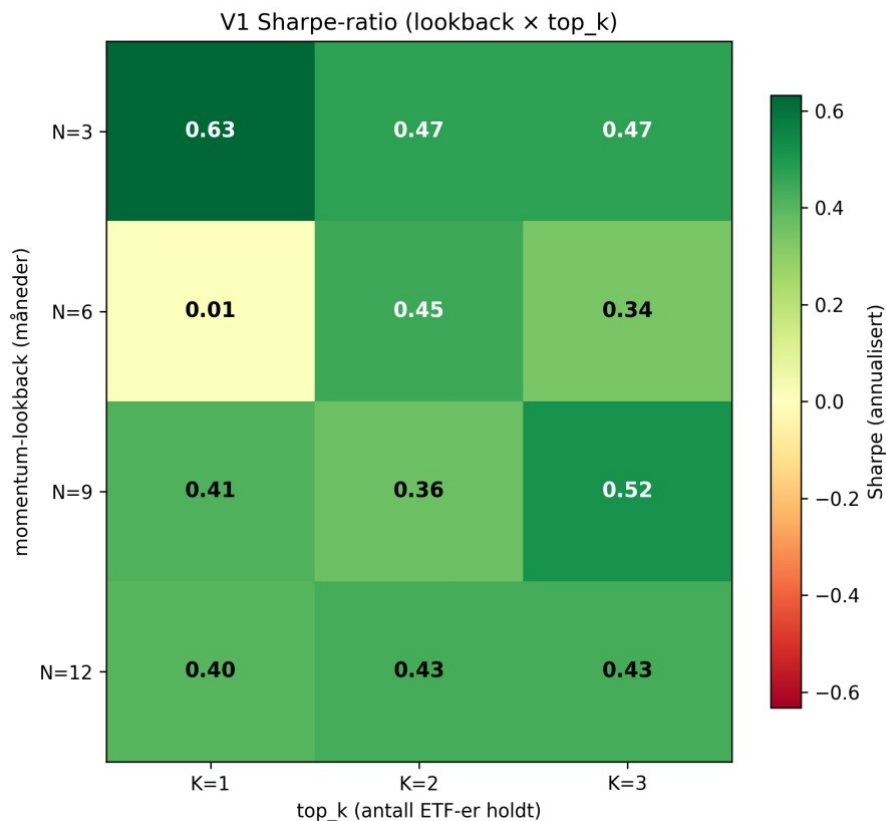
Resultater

V1 oppfyller alle tre kriterier: 12 av 12 har positiv Sharpe, standardavvik (0.148) er mindre enn snitt (0.411), og verste max DD (-33.7 %) er innenfor terskelen (-40.2 %). Konklusjon: V1 er ROBUST per definisjonen.

Hovedtabell for parametersensitivitet (Sharpe ratio):

Lookback	K=1	K=2	K=3
N=3 mnd	0.63	0.47	0.47
N=6 mnd	0.01	0.45	0.34
N=9 mnd	0.41	0.36	0.52
N=12 mnd	0.40	0.43	0.43

Visualiseringen nedenfor (Figur 1) bekrefter mønsteret i tabellen og illustrerer den spesifikke fragiliteten ved N=6, K=1. Heatmapen viser at de fleste parameterkombinasjoner gir Sharpe-verdier mellom 0.34 og 0.63, med et tydelig avvik der N=6 kombineres med konsentrert posisjon (K=1).



Figur 1: V1 Sharpe-ratio på tvers av parameterkombinasjoner. Heatmapen viser tydelig at de fleste kombinasjoner gir konsistente resultater i området 0.34-0.63, men at N=6, K=1 framstår som en isolert svak celle (Sharpe 0.01). Resultatet ble verifisert som ekte fragilitet ved konsentrert posisjon, ikke en implementeringsbug.

Snitt Sharpe på tvers av kombinasjoner: 0.41. Til sammenligning er benchmark IMEU Sharpe 0.45 og 60/40 Sharpe 0.42.

In-sample / out-of-sample analyse

Som tilleggssanalyse for robusthet ble parametergridet kjørt på to halvdelar av perioden separat: 2018-2021 (IS) og 2022-2024 (OOS). Korrelasjonen mellom IS-Sharpe og OOS-Sharpe på tvers av de 12 parameterkombinasjonene er kun 0.25, noe som indikerer at det var liten til ingen pålitelig sammenheng mellom hvilke parametre som fungerte best i første halvdel og hvilke som fungerte best i andre halvdel.

Konkret hadde vårt opprinnelige default-valg (N=6, K=2) en IS-Sharpe på 0.82, men en OOS-Sharpe på -0.09. Hvis vi på slutten av 2021 hadde valgt de parametrene som fungerte best, ville vi ha valgt en kombinasjon som ikke fungerte i de påfølgende tre årene.

Nyansering av robusthetsfunnet

Det formelle resultatet (ROBUST per definisjonen) gir en svakere konklusjon enn det første øyekast tilsier. Tre observasjoner kvalifiserer betydelig hva 'robust' egentlig betyr:

Edgen sitter i drawdown, ikke avkastning. 12 av 12 V1-varianter har lavere max drawdown enn buy-and-hold, men kun 4 av 12 slår samme benchmark på Sharpe ratio. Gjennomsnittlig CAGR (4.4 %) er lavere enn benchmark (6.33 %). V1 reduserer risiko ved å gi fra seg avkastning.

Out-of-sample-instabiliteten er det mest avslørende funnet. IS-OOS Sharpe-korrelasjon på 0.25 indikerer at strategien ikke har en stabil parameter-konfigurasjon over tid. Vi har ingen pålitelig grunn til å tro at de parametrene som fungerte i fortiden ville fungere i fremtiden.

Vår default-konfigurasjon (N=6, K=2) var moderat flaks. På full periode er den grei, men N=6-raden har lavest snitt-Sharpe (0.26 på tvers av top-K). N=3 hadde høyest snitt (0.52). Den konvensjonelle 6-måneders momentum-horisonten var altså dårligst i snitt i vårt utvalg.

Det ærlige funnet er at V1 er teknisk robust per definisjonen, men robustheten betyr 'kollapser ikke på tvers av parametre', ikke 'har påpekt edge over benchmark'.

4.5 Absolutt momentum baseline

Etter at V1-parameterrobusthet hadde indikert at V1-edgen var marginal og ustabil, ble den enkleste mulige trendfølgingsstrategien testet som baseline. Dersom en simpel switch mellom bred indeks og statsobligasjon kunne matche eller slå V1, ville det bekrefte at kompleksiteten i V1 (fem faktor-ETF-er med månedlig rotasjon) ikke la til verdi utover den underliggende trendfølgingsmekanismen.

Strategilogikk

Strategien følger Faber (2007) i sin enkleste form. Hver måneds siste handelsdag:

- Hvis IMEU close > SMA(N): hold 100 % IMEU
- Hvis IMEU close <= SMA(N): hold 100 % IEGA (europeiske statsobligasjoner)

To varianter ble testet:

- Variant A: SMA-periode 200 dager (klassisk Faber)
- Variant B: SMA-periode 252 dager (cirka 12 måneder)

Implementeringen følger samme arkitektur som faktorrotasjonsstrategiene og benytter samme motor, samme kostnadsmodell og samme cached data. V1 ble re-kjørt i samme runde for å sikre identisk sammenligningsgrunnlag, og endte på eksakt +41.36 % som referansen (konsistens-sjekken bestått).

Resultater

Hovedresultatene over perioden 2018-2024 er presentert i tabellen nedenfor:

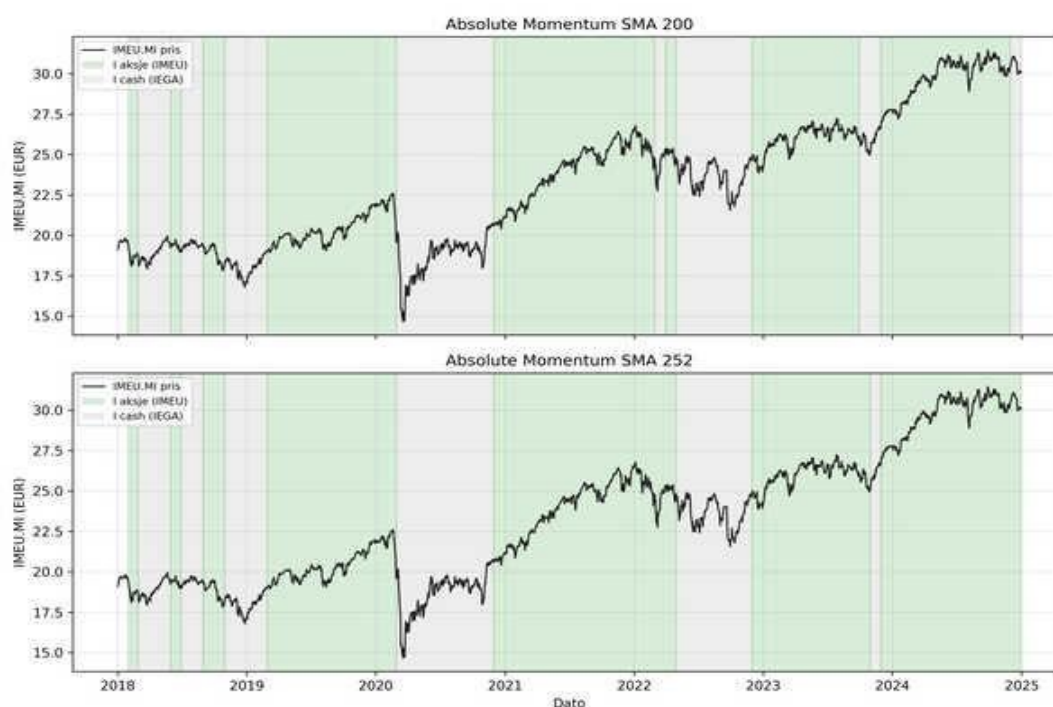
Strategi	Total avk.	CAGR	Sharpe	Sortino	Max DD	Calmar	Kostnad
AbsMom SMA 200	+26.4 %	3.40 %	0.35	0.42	-22.7 %	0.15	EUR 3 019
AbsMom SMA 252	+29.0 %	3.70 %	0.37	0.44	-18.0 %	0.21	EUR 2 396
Buy-and-hold IMEU	+53.7 %	6.33 %	0.45	0.53	-35.2 %	0.18	-
60/40 (statisk)	+29.8 %	3.80 %	0.42	0.50	-24.1 %	0.16	-

V1 ren momentum	+41.4 %	5.07 %	0.45	0.57	-20.7 %	0.24	EUR 4 917
-----------------	---------	--------	------	------	---------	------	-----------

Analyse av regimeskift

Et sentralt funn er hvordan strategiene håndterte COVID-krakket i mars 2020. Begge varianter gikk til cash den 28.02.2020, noe som riktignok dempet det innledende fallet. Imidlertid forble strategien i cash gjennom hele V-formede opphenting som fulgte. Variant A re-engasjerte i markedet 30.11.2020, og Variant B enda senere. På det tidspunktet var det meste av rallyet fra mars-bunnen allerede over - strategien gikk dermed glipp av cirka 40 % avkastning fra bunnpunktet.

Figur 2 nedenfor illustrerer allokeringshistorikken for begge varianter visuelt. Grønn bakgrunn indikerer eksponering mot aksjer (IMEU), grå bakgrunn indikerer cash-posisjon (IEGA). Den sammenhengende grå blokken i 2020 viser COVID-cash-perioden tydelig, og den lengre grå blokken i 2022 viser bear-market-perioden. Forskjellen mellom SMA 200 og SMA 252 er også synlig. Kortere SMA gir flere whipsaw-effekter (raske regimeskift uten klar retning), spesielt i 2018.



Figur 2: Allokeringshistorikk for Absolutt Momentum SMA 200 (over) og SMA 252 (under). Grønn bakgrunn = posisjon i IMEU, grå bakgrunn = posisjon i IEGA (cash). Den sorte linjen viser IMEU-prisen. Lange cash-perioder gjennom 2020 (COVID) og 2022 (bear) illustrerer strategiens tendens til å gå glipp av V-formede opphenter.

Dette er en klassisk svakhet ved trendfølgingsstrategier som rebalanserer månedlig. Signalene er for trege til å fange opp V-formede krakk. Strategien fanger en del av nedsiden, men misser opphenteren. I langvarige nedturer (slik som 2008 eller 2000-2002 dot-com-krasjet) ville strategien sannsynligvis fungert bedre, men disse periodene er ikke en del av vårt datautvalg.

I 2022 viste variantene ulik atferd. SMA 252 (lengre vindu, tregere signaler) holdt seg konsekvent i cash gjennom hele bear-markedet og unngikk dermed mye av nedsiden. SMA 200 var derimot utsatt for whipsaw-effekter, der den vekslet mellom cash og aksjer flere ganger uten klar retning.

Kostnadsanalyse

Absolutt momentum-strategiene har betydelig lavere transaksjonskostnader enn V1, som var forventet. Variant A påførte total kostnad på EUR 3 019 og Variant B på EUR 2 396 over perioden, mot V1 sine EUR 4 917. Variant B er dermed omtrent halvparten så kostbar som V1, men leverer betydelig lavere avkastning.

Sammenligning mot V1

Resultatet er ubehagelig men ærlig: V1-kompleksiteten slår faktisk den simple switchen. V1 ga +41.4 % mens AbsMom-variantene ga 26-29 %. Sharpe-ratioene er også høyere for V1 (0.45 vs 0.35-0.37). V1 taper kun mot SMA 252 på max drawdown (-20.7 % vs -18.0 %). Denne sammenligningen må samtidig leses sammen med funnet i kapittel 4.4: V1 slår den enkle switchen på full periode, men uten en parameterkonfigurasjon som er stabil ut av utvalget. De to funnene står ikke i motsetning, men de begrenser hverandre, og avstemmes i kapittel 4.8.

Dette nyanserer en viktig observasjon. Det er ikke trendfølging i seg selv som er problemet. Det er trendfølging på indekssnivå (Absolute Momentum) som er svakere enn trendfølging på faktornivå (V1 ren momentum). V1 holder alltid en posisjon i de mest attraktive faktor-ETF-ene, mens Absolute Momentum går helt til cash og dermed misser opphentingene i V-krakk.

Sammenligning mot statisk 60/40

Det viktigste enkeltfunnet fra denne testen er sammenligningen mot 60/40-porteføljen. En statisk 60/40-allokering (60 % IMEU, 40 % IEGA, rebalansert månedlig) leverte +29.8 % med Sharpe 0.42 og max DD -24.1 %.

60/40 slår dermed begge AbsMom-variantene på:

- Total avkastning (29.8 % vs 26.4 / 29.0 %)
- Sharpe ratio (0.42 vs 0.35 / 0.37)
- Sortino ratio (0.50 vs 0.42 / 0.44)

Og dette uten timing, uten månedlige regimebeslutninger, og uten transaksjonskostnader knyttet til regimeskift. En kjedelig statisk miks slår dermed begge tidsstyrte variantene på nesten alle relevante mål. Det eneste 60/40 taper på er max drawdown, der SMA 252 har lavest av alle (-18.0 % vs -24.1 %).

Sammenligningen er entydig. Hele SMA-apparatet, med all sin teoretiske appell og kompleksitet, klarte ikke å tjene inn sin egen eksistens i denne perioden. Hvis HULC AS ønsker lavere drawdown enn buy-and-hold, dominerer en statisk 60/40 begge SMA-variantene på nesten alle mål.

Forbehold

Det er viktig å fremheve at 7 år med ett V-formet krakk er strukturelt det verste regimet for månedlig trendfølging. Faber-strategiens historiske rykte ble tjent i 2008 og 2000-2002, som var langvarige nedturer der månedlig rebalansering var rask nok til å fange det meste av nedsiden uten å misse oppsiden. Slike perioder er ikke en del av vårt datautvalg.

Konklusjonen om Absolutt momentum-strategien er derfor suggestiv og ikke konklusiv. Strategien fungerer dårlig i V-formede krakk, men kan potensielt fungere bedre i langvarige nedturer. 60/40-sammenligningen er imidlertid uavhengig av krakk-formen, og denne sammenligningen taler tydelig mot å ta i bruk Absolutt momentum-strategien.

Konklusjon for absolutt momentum

Anbefalingen er å ikke ta i bruk noen av AbsMom-variantene for HULC AS. Ønsker selskapet lavere drawdown enn buy-and-hold, dominerer en statistisk 60/40-portefølje dem på nesten alle mål. Samlet på tvers av faktorrotasjonen og absolutt momentum finner vi at verken intraday mean-reversion, faktorrotasjon eller indeks-trend har en pålitelig edge på dette utvalget, og en kjedelig statistisk allokering matcher eller slår dem alle.

4.6 Cross-sectional momentum på amerikanske enkeltaksjer

Cross-sectional momentum på 12 måneder eksklusive siste måned, topp 15 likevektet, halvårlig rebalansering. S&P 500-medlemmer med historisk medlemskap, 55 rebalanseringer fra 30. juni 1999 til 30. juni 2026. Kostnader, T+1-oppgjør og 22 % skatt som beskrevet i kapittel 3.1. Referansen er SPY kjøp-og-hold med samme kostnadsbehandling for den initiale posisjonen. Strategien er tatt med som kontrollgruppe. Den viser hva momentum alene gir på universet, slik at multifaktorstrategiens eventuelle bidrag utover pris kan isoleres i kapittel 4.8.

	Test A 50k	Test A 100k	S&P 500
Total avkastning	214,58 %	248,74 %	803,25 %
CAGR	4,32 %	4,72 %	8,46 %
Sharpe	0,293	0,305	0,519
Sortino	0,378	0,394	0,663
Maks drawdown	-86,23 %	-85,64 %	-55,19 %
Calmar	0,050	0,055	0,153
Treffrate	48,00 %	48,52 %	—
Turnover, årlig (tosidig)	284,26 %	284,26 %	0 %
Betalt skatt (USD)	3 075	7 190	0
Sluttverdi (USD)	19 929	44 240	57 174 / 114 492

Startkapitalen er 6 372 USD ved 50 000 NOK og 12 743 USD ved 100 000 NOK, omregnet med 7,8471. Referansen slås ikke på noe måltall. Strategien ligger 4,1 prosentpoeng under indeksen i årlig avkastning, med under halvparten så høy Sharpe og en drawdown 31 prosentpoeng dypere.

Kostnadene er dekomponert per komponent. Kommisjonen er nesten identisk i kroner på de to kapitalnivåene, 1 313 mot 1 367 dollar, fordi antall ordrer er nesten likt og gulvet på én dollar binder i begge. Den utgjør dermed dobbelt så stor andel av omsetningen på det laveste nivået, 0,296 mot 0,142

% . Samlet kostnadstrekk er 1,53 prosentpoeng i året ved 50 000 NOK og 1,10 ved 100 000. Det er kommisjonsgulvet målt direkte, og det er grunnlaget for vurderingen i kapittel 3.1.3.

Kostnad	50k, USD	50k, pp/år	100k, USD	100k, pp/år
Kommisjon	1 313	0,840	1 367	0,404
Spread	89	0,057	193	0,057
Slippage	994	0,636	2 150	0,635
Sum	2 396	1,533	3 710	1,095

Rangeringsbåndet, altså at en eid posisjon beholdes så lenge den ligger innenfor topp 23, selv om porteføljen bare holder 15 navn, gir 0,15 prosentpoeng høyere avkastning ved 50 000 NOK og 0,24 ved 100 000. Turnover faller fra 313 til 284 %, og 63 posisjoner ble beholdt utelukkende fordi de lå innenfor bufferen. Verdt å merke seg er at båndet her øker skatten, fra 2 625 til 3 075 dollar, samtidig som det øker avkastningen. Det er motsatt av mekanismen man skulle vente. En nærliggende forklaring er at varianten uten bånd realiserer flere tap, som ikke utløser skatt, men mekanismen er ikke målt og skal ikke framstilles som om den var det.

Fordelingen over femårs delperioder er den samme historien fortalt seks ganger.

Periode	Test A	S&P 500	Meravkastning
1999–2004	–25,51 %	–10,57 %	–14,95 %
2004–2009	–52,72 %	–11,05 %	–41,66 %
2009–2014	+101,05 %	+135,58 %	–34,53 %
2014–2019	+17,79 %	+65,31 %	–47,52 %
2019–2024	+102,65 %	+99,15 %	+3,50 %
2024–2026 *	+86,93 %	+44,79 %	+42,14 %

* Ufullstendig delperiode på 2,09 år. Den er tatt med framfor utelatt, siden en utelatt delperiode er en stille utvalgsbeslutning.

Strategien slår referansen i 2 av 6 delperioder, og begge er de to siste. Det er verdt å merke seg mot totalbildet. Den strategien som taper klart over hele historikken, er den som har vært best de siste årene.

NAV-banen forklarer hvorfor avkastningen er lav til tross for at strategien i perioder gjør det svært godt. Porteføljen steg fra 6 335 til 13 240 dollar innen 23. mars 2000, altså en dobling på ni måneder, og falt deretter. Bunnen kom 20. november 2008 med 86,2 % under toppen, altså 1 824 dollar. Den gamle toppen ble ikke tatt igjen før 27. juni 2025, over 25 år senere.

Stressepisode	Vindu	Test A	S&P 500
Dot-com	2000-03 → 2002-12	–74,5 %	–41,6 %
Finanskrisen	2007-10 → 2009-03	–75,0 %	–54,6 %
Covid	2020-02 → 2020-04	–35,5 %	–30,8 %

Beholdningene bekrefter at strategien gjorde nøyaktig det den er bygget for. I juni 2000 eide den AMD, Sun, Micron, Oracle og Siebel, altså dot-com-navnene rett før fallet. I juni 2008 eide den energi og kull rett før råvarekollapsen. I desember 2020 eide den vekstnavn rett før korreksjonen i 2021. Momentum kjøper det som har steget, og konsentrasjonen på 15 navn gjør hver av de tre episodene til et fullt sammenbrudd framfor et tilbakeslag. Mønsteret er kjent i litteraturen som momentumkollaps (Daniel og Moskowitz, 2016). Resultatet er ikke at strategien er feilimplementert, men at strategien er det den er.

Skatten er lav i absolutte tall, 3 075 dollar ved 50 000 NOK. Det er ingen fordel. Strategien realiserte store tap i 2001, 2002 og 2008, og framførte dem i 23 av 28 år. Lav skatt er her konsekvensen av å tape penger.

4.7 Multifaktor (QVM) på amerikanske enkeltaksjer

Multifaktorstrategien rangerer på en sammensatt score av kvalitet, verdi og momentum med lik vekt på de tre blokkene, innenfor sektorgruppe, slik konstruksjonen er beskrevet i kapittel 3.2.3. Topp 15 likevektet, halvårlig rebalansering, 55 rebalanseringer fra 30. juni 1999 til 30. juni 2026. Alt annet er identisk med Test A. Samme univers, samme kostnader, samme oppgjør, samme skatt og samme rangeringsbånd. Konfigurasjonen importeres fra Test A framfor å gjentas, slik at de to ikke kan komme i utakt.

Kriteriene ble definert på forhånd, før testen ble kjørt. Strategien må slå både den passive referansen og cross-sectional momentum etter kostnader og skatt, og ha statistisk signifikant alfa i faktorregresjonen. Resultatet måles mot dem til slutt i seksjonen.

	Test B 50k	Test B 100k	S&P 500
Total avkastning	436,82 %	458,22 %	803,25 %
CAGR	6,40 %	6,55 %	8,46 %
Sharpe	0,394	0,400	0,519
Sortino	0,517	0,526	0,663
Maks drawdown	-58,00 %	-57,96 %	-55,19 %
Calmar	0,110	0,113	0,153
Treffrate	56,97 %	56,97 %	—
Turnover, årlig (tosidig)	251,66 %	251,80 %	0 %
Betalt skatt (USD)	6 934	14 667	0
Sluttverdi (USD)	34 027	70 850	57 205 / 114 554

Referansen slås ikke. Multifaktorstrategien ligger 2,1 prosentpoeng under S&P 500 i årlig avkastning, med lavere Sharpe, lavere Calmar og marginalt dypere drawdown. Det er hovedresultatet. Resten av seksjonen forklarer hvordan det oppsto, ikke hvorfor det egentlig var bedre enn det ser ut.

Kostnadstrekket er 0,75 prosentpoeng i året ved 50 000 NOK og 0,61 ved 100 000, altså under halvparten av Test A sitt på tross av nesten samme turnover. Grunnen er at strategien har mer kapital å handle på gjennom perioden. Den taper seg ikke ned til under 2 000 dollar underveis, og kommisjonsgulvet biter dermed langt mindre. Kostnadsnivået er altså like mye en funksjon av hvordan strategien har gått som av hvordan den handler.

Kostnad	50k, USD	50k, pp/år	100k, USD	100k, pp/år
Kommisjon	1 214	0,272	1 279	0,137
Spread	222	0,050	466	0,050
Slippage	1 886	0,423	3 962	0,424
Sum	3 321	0,745	5 707	0,611

Rangeringsbåndet gir 0,45 prosentpoeng høyere avkastning ved 50 000 NOK og 0,43 ved 100 000, altså samme størrelsesorden og samme fortegn som i Test A. 72 posisjoner ble beholdt utelukkende fordi de lå innenfor bufferen. Dette tallet er en rettelse. Med den tidligere motoren, som hoppet over 16 av 55 rebalanseringer, ble båndet målt til minus 0,12 prosentpoeng, og rapporten konkluderte med at det gjorde det motsatte av i Test A. Med alle rebalanseringene på plass virker det som forventet. Poenget generaliserer utover denne tabellen. Et funn om en driftmekanisme er også et funn om hvor ofte mekanismen får virke.

Fordelingen over delperioder er det mest urovekkende i hele resultatet.

Periode	Test B	S&P 500	Meravkastning
1999–2004	+58,18 %	–10,57 %	+68,75 %
2004–2009	–4,19 %	–11,05 %	+6,86 %
2009–2014	+106,52 %	+135,58 %	–29,06 %
2014–2019	+6,41 %	+65,31 %	–58,90 %
2019–2024	+25,99 %	+99,15 %	–73,16 %
2024–2026 *	+29,22 %	+44,79 %	–15,56 %

* Ufullstendig delperiode på 2,09 år.

Strategien slår referansen i 2 av 6 delperioder, og de to er de to første. Mønsteret er tydelig. Hele meravkastningen ble tjent før 2009, og hver eneste delperiode etter finanskrisen taper mot indeksen. En strategi som virket i de to første tredjedelene av utvalget og ikke i den siste, er ikke en strategi med edge. Den er en strategi med en periode.

NAV-banen går fra 6 339 til 34 027 dollar, med toppen 34 503 dollar den 28. juli 2026. Verste drawdown er 58,00 %, fra 18 293 dollar 13. juli 2007 til 7 684 dollar 3. mars 2009, og den gamle toppen ble ikke tatt igjen før 19. mars 2014, altså seks og et halvt år under vann.

Stressepisode	Vindu	Test B	S&P 500
Dot-com	2000-03 → 2002-12	–5,5 %	–41,6 %

Finanskrisen	2007-10 → 2009-03	-57,1 %	-54,6 %
Covid	2020-02 → 2020-04	-45,6 %	-30,8 %

De tre episodene forklares av beholdningene, og de peker i hver sin retning. I juni 2000 eide porteføljen jernbane, kjemi, forsikring, industri og en avis, ikke dot-com-navnene. Det er verdi- og kvalitetsblokkene som overstyrer momentum. Et selskap uten inntjening får negativ fortjeneste mot markedsverdi og faller til bunns i verdiblokka uansett hvor sterkt kursen har steget. Dot-com er den ene episoden strategien håndterte godt, og den forklarer nesten hele meravkastningen over 27 år.

I juni 2008 eide den derimot energi og råvarer sammen med fire forsikringsselskaper, altså selskaper som var billige på regnskapstall rett før både råvarekollapsen og finanskrisen traff nettopp de sektorene. Verdiblokka pekte inn i krisen i stedet for bort fra den, og drawdownen ble dypere enn indeksens. I desember 2020 eide den boligbyggere, forsikring og helse, falt 46 % mot indeksens 31 % i covid-krakket, og hentet det ikke inn igjen i de påfølgende årene.

Skatten er 6 934 dollar ved 50 000 NOK, altså dobbelt så høy som i Test A i absolutte tall. Grunnen er enkel. Strategien tjente mer penger. Test A hadde framført tap i 23 av 28 år, Test B i 10. Det er ikke skatteoptimalisering, det er at strategien taper sjeldnere.

Kvalifiseringen er det ene steget Test B har som Test A ikke har. Et selskap må ha ferske nok verdier i alle tolv fundamentalfeltene. Fordelingen over de 55 rebalanseringene viser at kravet i praksis er mildt.

Steg	Median	Min	Maks
Indeksmedlemmer	500	500	505
Kvalifiserte	496	466	500
Ekskludert: manglende felt	5	0	20
Ekskludert: data eldre enn 180 dager	1	0	16
Ekskludert av strategien	1	0	4
Rangert	495	466	500

Sektorinndelingen ga tolv grupper ved hver eneste rebalansering, så mekanismen for sammenslåing av små grupper ble aldri utløst. De 64 strategieksklusjonene til sammen er selskaper der en hel blokk manglet, i praksis negativ egenkapital eller negativ enterprise value. Porteføljen brukte 310 unike navn over de 55 rebalanseringene. Survivorship-skjevheten som følger av eksklusjonene er behandlet i kapittel 3.2.2.

Det siste funnet gjelder konstruksjonen selv, og det ble først synlig da scoren ble målt. Medianscoren blant de valgte navnene er systematisk høyest i momentumblokka.

Blokk	Snitt	Standardavvik	Min	Maks
Kvalitet	0,213	0,056	0,051	0,329
Verdi	0,239	0,057	0,103	0,335

Momentum	0,389	0,037	0,278	0,450
----------	-------	-------	-------	-------

Årsaken er ikke at momentum er vektet høyere. Vektene er like. Årsaken er at momentum er ett enkelt mål mens kvalitet og verdi er gjennomsnitt av tre og fire mål hver. Et gjennomsnitt av flere rangeringer trekker mot midten, så en blokk med flere mål får mindre spredning enn en blokk med ett. Målt på tvers av alle rangerte navn er hvert enkeltmål like spredt, med standardavvik 0,289, mens blokkene ikke er det, omtrent 0,19 for kvalitet, 0,20 for verdi og 0,289 for momentum. Grovt regnet på spredningen alene tilsvarer det en effektiv vekt på omtrent 42 % til momentum mot 29 % hver til kvalitet og verdi.

Lik vekt i koden er altså ikke lik vekt i porteføljen, og ingen har bestemt at den ikke skulle være det. Dette er en konstruksjonsfeil, og den er verdt å teste i robusthetsrunden framfor å rette i etterkant av et resultat. En normalisering av blokkscorene før vekting ville gjort vektene reelle. Flagget som gjør det finnes i koden, men det er lagt inn etter at testen var kjørt, og står derfor av som standard. Å endre konstruksjonen etter å ha sett resultatet ville gjort robusthetsrunden om til en etterrasjonalisering.

Målt mot kriteriene slår strategien cross-sectional momentum på hvert eneste måltall, den slår ikke den passive referansen på noe måltall, og faktorregresjonen er ikke kjørt. Det ene kriteriet som er avgjort i strategiens disfavør er det som veier tyngst. En strategi som taper mot et indeksfond etter kostnader og skatt, over 27 år, kan ikke forsvares med at den slo en annen aktiv strategi. Faktorregresjonen gjenstår og vil kunne vise om det som er igjen av avkastning er eksponering mot kjente faktorer, men den kan ikke endre hovedresultatet.

Robusthet og ut-av-utvalg

Et enkeltresultat sier lite om en strategi er noe verdt. Spørsmålet er om resultatet er en egenskap ved konstruksjonen eller ved det parametervalget som tilfeldigvis ble kjørt. Det er testet med et sveip over 30 kombinasjoner. Fem vektsett, altså lik vekt, kun kvalitet, kun verdi, kun momentum og kvalitet pluss verdi, kombinert med tre porteføljestørrelser på 10, 15 og 25 navn, med og uten sektornøytralisering. Samme periode, samme kostnader, samme skatt og samme oppgjør som i testene over. Målene beregnes én gang per rebalanseringsdato og scores på nytt per kombinasjon, slik at fundamentaloppslaget er identisk på tvers av rutenettet.

Kriteriene ble skrevet i koden før første kjøring og er ikke endret etterpå. Robust betyr her at strategien ikke kollapser i parameterrommet, ikke at den har edge. Formuleringen er arvet fra sveipet i kapittel 4.4 og er like nødvendig her.

Kriterium	Måling	Utfall
Minst 75 % har positiv Sharpe	30 av 30	Oppfylt
Standardavvik lavere enn snitt	0,065 mot 0,422	Oppfylt
Ingen drawdown mer enn 5 pp verre enn referansen	Verste -84,89 %	Ikke oppfylt
Minst 75 % med lavere volatilitet enn Test A	29 av 30	Oppfylt

Tre av fire kriterier er innfridd. Det tredje er ikke, og ikke i et hjørne av rutenettet. Halve rutenettet, 15 av 30 kombinasjoner, har en drawdown verre enn terskelen på 60,19 %. De verste er momentum alene uten sektornøytralisering, altså i praksis Test A, men også alle seks verdi-variantene, tre av seks kvalitet-varianter og to med lik vekt bryter grensen. Dybden på fallene er ikke en egenskap ved ett uheldig parametervalg. Samlet dom er at rutenettet ikke er robust.

Volatilitetstallene i denne tabellen er annualisert fra daglige avkastninger. Det samme gjelder dekomponeringstabellene sist i kapittel 4.8. Tabellen over volatilitetsdrag tidlig i kapittel 4.8 annualiserer derimot fra månedlige avkastninger, og oppgir derfor lavere tall for de samme strategiene. Begge er gyldige, og den daglige er høyere fordi den fanger dagsvingninger som utjevnes over en måned. Sammenligningen i det fjerde kriteriet bruker samme estimator på begge sider.

Et utdrag av rutenettet, sortert på Sharpe, viser mønsteret. Referansen har i samme periode CAGR 8,46 %, Sharpe 0,519 og maks drawdown 55,19 %.

#	Vekter	K	Sektor	CAGR	Sharpe	Maks DD	IS	OOS
1	Q+V	25	flat	9,27 %	0,514	-49,59 %	0,569	0,451
2	Kun Q	25	flat	8,84 %	0,505	-52,36 %	0,351	0,674
3	Q+V	25	sektor	8,50 %	0,491	-56,18 %	0,480	0,501
4	Kun Q	15	flat	8,88 %	0,490	-56,13 %	0,371	0,637
5	Q+V	15	sektor	8,77 %	0,490	-54,88 %	0,474	0,510
6	Kun V	15	flat	9,84 %	0,490	-71,99 %	0,406	0,583
7	Q+V	10	sektor	9,03 %	0,486	-51,09 %	0,523	0,443
20	Lik (Test B)	15	sektor	6,40 %	0,394	-58,00 %	0,315	0,478
29	Kun M	15	flat	4,69 %	0,304	-84,89 %	0,007	0,674
30	Kun M	10	sektor	4,47 %	0,300	-61,56 %	-0,044	0,657

To mønstre går igjen. Flere navn er bedre: 25 navn gir høyest gjennomsnittlig Sharpe i alle fem vektsettene og laveste drawdown i fire av fem, og rekkefølgen 25 foran 15 foran 10 er streng i fire av fem. Konsentrasjonen på femten navn koster, og den koster mest i volatilitet. Og momentumblokka er den svakeste. De sju øverste kombinasjonene har ingen momentumblokk i det hele tatt. Først på plass åtte kommer en kombinasjon med momentum.

Den kombinasjonen Test B faktisk kjørte, ligger på plass 20 av 30. Den er ikke et heldig punkt i rutenettet, den er et middelmådig et. Nitten kombinasjoner gjorde det bedre, og de deler to trekk, mindre momentum og flere navn.

Det er fristende å lese den beste ruten som en oppdagelse. Kvalitet pluss verdi med 25 navn uten sektornøytralisering gir 9,27 % årlig avkastning mot referansens 8,46 %, med en drawdown på 49,6 % mot 55,2 %, og er dermed den eneste konfigurasjonen i hele rapporten som slår indeksen. Den lesningen holder ikke, og sveipets eget tall sier hvorfor.

Delingen i in-sample og out-of-sample går ved årsskiftet 2012/2013, på NAV-banen fra én sammenhengende kjøring framfor to separate backtester. Korrelasjonen mellom Sharpe i første og andre halvdel, over rutenettet, er minus 0,72. Til sammenligning var den pluss 0,25 for faktorrotasjonen i kapittel 4.4, og den ble den gang kalt det sterkeste enkeltbeviset mot den strategien. Hvilke parametre som virket i første halvdel sa nesten ingenting om hvilke som virket i andre. Minus 0,72 er ikke ingenting. Det er verre. Parametrene som gjorde det best i første halvdel gjorde det systematisk dårligst i andre.

Mekanismen er synlig i tabellen og er ikke støy. Momentum alene har Sharpe rundt null i første halvdel, fra minus 0,04 til 0,13, og mellom 0,66 og 0,76 i andre. Verdi alene og kvalitet pluss verdi har det motsatte mønsteret. Verdi og kvalitet ledet fra 1999 til 2012, momentum ledet fra 2013 til 2026. Rutenettet måler altså i stor grad hvilket faktorregime hver halvdel tilhørte, ikke hvilke parametre som er gode. Konsekvensen er konkret. Hadde man valgt vektsett på data fram til 2012, ville man valgt kvalitet pluss verdi, og fått den svakeste halvdelen av utfallene etterpå. Den beste kombinasjonen in-sample rangerer nummer 29 av 30 out-of-sample.

Sveipet svarer samtidig på spørsmålet kapittel 4.8 lot stå. Er den lave volatiliteten en egenskap ved konstruksjonen eller ved parametervalget? Det fjerde kriteriet svarer at den er en egenskap ved konstruksjonen. 29 av 30 kombinasjoner har lavere volatilitet enn ren momentum, og den ene som ikke har det er momentum alene med ti navn uten sektornøytralisering, altså den kombinasjonen som ligger nærmest Test A i utgangspunktet. Nettopp det unntaket er verdt å merke seg, for det peker på hvilken mekanisme som virker. Dekomponeringen sist i kapittel 4.8 viser at det i hovedsak er sektornøytraliseringen og ikke de fundamentale blokkene som senker volatiliteten.

Konstruksjonsfeilen med blokkvektene er også testet. Normalisering av blokkcorene før vektning, slik at like vekter blir like vekter, gir seks bedre og seks dårligere utfall av tolv sammenligninger, uten mønster i hvilke. Svakheten er ekte og uten praktisk konsekvens på dette rutenettet. Det er verdt å skrive ned nettopp fordi det motsatte var forventet.

Ingen enkeltkombinasjon skal forfremmes på dette grunnlaget. Kvalitet pluss verdi med 25 navn ser best ut, men den er valgt etter kjøring i et rutenett der utvalg på første halvdel beviselig peker feil vei. Å plukke den ville vært å gjøre nøyaktig det sveipet demonstrerer at man ikke skal gjøre.

4.8 Sammenfattende sammenligningstabell

Denne seksjonen svarer på spørsmålet rapporten ble skrevet for. Tilfører de fundamentale blokkene noe utover pris alene? Test A og Test B kjøres på nytt i samme prosess og på samme kurshistorikk framfor å sammenligne to sett resultatfiler, og referansen skaleres én gang og brukes for begge.

Det første spørsmålet må være om sammenligningen i det hele tatt måler to ulike porteføljer. Er utvalgene i praksis de samme navnene, sammenligner man to varianter av samme portefølje og kaller det to strategier. Det gjør man ikke her. De to porteføljene deler i snitt 2,15 av 15 navn, altså 14,3 %, med median 2 og spenn fra 0 til 5. Seks av 55 rebalanseringer har ikke ett eneste felles navn. Overlappet er dessuten stabilt gjennom hele historikken, mellom 1,95 og 2,50 i snitt per tiår.

Momentumblokka utgjør en tredjedel av multifaktorscoren, og et navn må være godt på kvalitet og verdi i tillegg for å komme inn. I snitt klarer to av femten det.

Test B slår Test A på hvert eneste måltall: 6,40 mot 4,32 % i årlig avkastning ved 50 000 NOK, 0,394 mot 0,293 i Sharpe, 28 prosentpoeng grunnere drawdown og lavere turnover. Ingen av dem slår referansen på noe måltall. Differansen lar seg dele i to ved å kjøre begge på nytt uten kostnader og uten skatt.

Kapital	A brutto	B brutto	A netto	B netto	Allokering	Friksjon
50 000	6,81 %	8,52 %	4,32 %	6,40 %	+1,71 pp	+0,37 pp
100 000	6,79 %	8,57 %	4,72 %	6,55 %	+1,77 pp	+0,06 pp

Åtte av ni tideler av forskjellen er allokering, altså hva strategiene eide. Friksjonen favoriserer også multifaktor, men marginalt. Lavere turnover sparer kostnader, mens høyere avkastning gir mer skatt, og de to utligner nesten hverandre. Verdt å merke seg for seg selv er at friksjonen koster Test A 2,5 prosentpoeng i året og Test B 2,1. Begge strategiene bruker altså rundt en fjerdedel av bruttoavkastningen på kostnader og skatt.

Neste tabell er den viktigste i rapporten, fordi den viser at differansen ikke er der man skulle tro.

	Snitt per mnd	Annualisert snitt	Annualisert vol	CAGR	Volatilitetsdrag
Test A 50k	0,64 %	7,68 %	26,13 %	4,32 %	3,36 pp
Test B 50k	0,66 %	7,89 %	18,21 %	6,40 %	1,49 pp
S&P 500	0,77 %	9,28 %	15,11 %	8,46 %	0,82 pp

Multifaktor tjener ikke mer enn ren momentum. Den taper mindre på veien. På 100 000-nivået er de aritmetiske gjennomsnittene 8,06 % for Test A og 8,04 % for Test B, altså identiske. Hele CAGR-differansen på 2,1 prosentpoeng er forskjellen i volatilitetsdrag, 26 % årlig volatilitet mot 18 %. Et tap på 50 % krever 100 % for å hentes inn, og Test A hadde tre slike episoder. Hva som skaper den lavere volatiliteten, er dekomponert i neste avsnitt, og svaret er ikke det man skulle tro.

Det gjør resultatet både bedre og dårligere enn det ser ut. Bedre, fordi lavere volatilitet er en ekte og verdifull egenskap som ikke forsvinner så lett som en meravkastning gjør. Dårligere, fordi det ikke finnes noe tegn til at de fundamentale blokkene finner selskaper som stiger mer. De finner selskaper som faller mindre, og indeksen gjør begge deler bedre enn begge strategiene.

Spørsmålet er så om forskjellen er større enn støy. En parvis test på månedlig avkastning over 326 observasjoner gir svaret.

Par	Korrelasjon	Snitt per mnd	t	p
B – A (50k)	0,648	+0,02 %	0,05	0,96
B – A (100k)	0,646	–0,00 %	–0,01	1,00
B – S&P 500 (50k)	0,808	–0,12 %	–0,67	0,50

A – S&P 500 (50k)	0,709	–0,13 %	–0,44	0,66
-------------------	-------	---------	-------	------

Den månedlige meravkastningen til Test B over Test A er ikke skillbar fra null. Det er ikke en marginal p-verdi. Den er 0,96, altså så nær null effekt som en test kan komme. Det er ingen selvmotsigelse mot CAGR-differansen på 2,1 prosentpoeng. Testen måler det aritmetiske snittet, og der er de like. Differansen i sluttverdi kommer fra andremomentet, ikke fra førstementet, og en t-test på snittet ser den ikke.

At heller ikke Test B mot referansen er signifikant, med $p = 0,50$, sier noe om hva 27 år faktisk rekker til. Med et månedlig standardavvik på 3 % i differansen kreves langt mer enn 326 observasjoner for å påvise en meravkastning på et par prosentpoeng. Fraværet av signifikans er ikke bevis på at strategiene er like. Det er en påminnelse om at dette utvalget ikke kan avgjøre spørsmålet i noen retning.

Beholdningene i de tre stressepisodene viser hvor lite de to strategiene har til felles. I juni 2000 eide Test A halvlederne og nettverksnavnene rett før fallet, mens Test B eide jernbane, kjemi, forsikring og en avis. Ett felles navn. I juni 2008 og desember 2020 var det ingen felles navn i det hele tatt. I 2008 tapte begge, fordi Test B hadde kjøpt de andre billige navnene i samme krise. I 2020 eide Test A vekstnavn og Test B boligbyggere og helse, og der vant Test A klart de påfølgende årene.

Konklusjonen er firedelt. Multifaktor tilfører noe reelt. Porteføljene er forskjellige, og Test B er bedre enn Test A på alle måltall. Men bidraget er risikoreduksjon og ikke meravkastning, og differansen er ikke statistisk skillbar fra null på månedlige data. Fordelingen over tid er verre enn nivået. Test B vinner de to første femårsperiodene og taper de to siste, så fortegnet på hva multifaktor tilfører avhenger av hvilken tiårsperiode man spør om, og den perioden som avgjorde saken, dot-com, er én hendelse. Og risikoreduksjonen kommer i hovedsak fra sektornøytraliseringsen, ikke fra de fundamentale blokkene, som vist i dekomponeringen over.

Hvor forbedringen faktisk kommer fra

Sammenligningen over har en svakhet som må håndteres før den kan tolkes. Test A og Test B skiller seg på to ting samtidig, nemlig signalet og sektornøytraliseringsen, og hele differansen er så langt tilskrevet det første. Det er ikke gitt at det er riktig. En egen kjøring skiller de to leddene ved å legge til én mekanisme om gangen.

Volatilitetstallene i dette avsnittet er annualisert fra daglige avkastninger, jf. noten i kapittel 4.7, og er derfor høyere enn tallene i tabellen over volatilitetsdrag. Mellomvarianten er Test A med rangering innenfor sektorgruppe og ingenting annet endret. Samme medlemsliste, samme sammenslåing av aksjeklasser, ingen kvalifisering på fundamentaldata. Den er kjørt med samme kostnader, skatt og oppgjør som de to andre.

Steg	Volatilitet	Maks drawdown	Sharpe	CAGR
Ren momentum, uten sektorbånd	31,63 %	–86,23 %	0,293	4,32 %

Med sektorbånd	24,00 %	-58,23 %	0,329	5,11 %
Med kvalitet og verdi i tillegg	21,91 %	-58,00 %	0,394	6,40 %

Sektorbåndet alene tar volatiliteten ned 7,63 prosentpoeng og drawdownen opp 27,99 prosentpoeng. De fundamentale blokkene, lagt oppå det, tar volatiliteten ned ytterligere 2,09 prosentpoeng og drawdownen opp 0,24. Sektorbåndet står dermed for 99 % av forbedringen i drawdown og 79 % av forbedringen i volatilitet.

Det er en vesentlig korreksjon av det bildet resten av kapittelet tegner. Risikoreduksjonen som skiller multifaktor fra ren momentum kommer i hovedsak fra å spre porteføljen over bransjer, ikke fra å velge selskaper på regnskapstall. En konsentrert momentumportefølje på femten navn havner i én bransje om gangen, og det er dét som gir fallene på 86 prosent.

Spørsmålet blir da om det er den faktiske bransjeinndelingen som virker, eller bare det å fordele porteføljen over tolv grupper uansett hvilke. Det er testet ved å stokke gruppemerkelappene tilfeldig, med fem frø, og ellers kjøre identisk.

Variant	Volatilitet	Maks drawdown	Sharpe	CAGR
Uten sektorbånd	31,63 %	-86,23 %	0,293	4,32 %
Permuterte grupper, snitt av fem	29,37 %	-81,03 %	0,319	5,16 %
Ekte bransjeinndeling	24,00 %	-58,23 %	0,329	5,11 %

Tilfeldige grupper reproducerer 19 % av effekten på drawdown og 30 % av effekten på volatilitet. Fire femtedeler av fallreduksjonen krever altså at gruppene faktisk er bransjer. Spredning i seg selv forklarer en del, men ikke mesteparten, og det er et argument for at sektornøytraliseringsen gjør noe reelt og ikke bare deler porteføljen i flere biter.

Det tredje leddet er kvalifiseringskravet. Et selskap må ha ferske verdier i alle tolv fundamentalfeltene for i det hele tatt å bli rangert, og det fjerner selskaper fra universet før noe signal er beregnet. Det leddet lar seg skille fra selve scoren ved å kjøre momentum alene på det kvalifiserte universet, altså med filteret på og signalet av.

Steg	Hva som legges til	CAGR	Maks drawdown	Sharpe
A	—	4,32 %	-86,23 %	0,293
A2	Sektorbånd	5,11 %	-58,23 %	0,329
A2 kvalifisert	Fundamentaldata som filter	5,75 %	-58,61 %	0,353
B	Fundamentaldata som signal	6,40 %	-58,00 %	0,394

Av differansen på 2,08 prosentpoeng i årlig avkastning mellom de to testene kommer 0,79 fra sektorbåndet, 0,63 fra kvalifiseringsfilteret og 0,65 fra selve multifaktorscoren. Under en tredjedel av forspranget skyldes altså det signalet hypotesen handler om.

Filterleddet er dessuten ikke en nøytral avgrensning, og retningen på skjevheten er allerede målt i kapittel 3.2.2. Selskaper som mangler fundamental dekning er overrepresentert blant dem som forsvant fra indeksen, med en relativ risiko på 2,73. Filteret hindrer strategien i å eie dem. De 0,63 prosentpoengene fra filterleddet er derfor ikke rene resultater, men delvis en måling av en kjent skjevhet med kjent retning. Størrelsesordenen på skjevheten er liten, som vist i samme kapittel, men leddet kan ikke leses som en egenskap ved strategien.

Dekomponeringen har en praktisk konsekvens som er verdt mer enn resten av sammenligningen. Sektornøytralisering krever ingen fundamentaldata. Den krever bare en bransjeklassifisering, som er tilgjengelig gratis, og den leverer nesten hele fallreduksjonen. Det kommersielle datasettet, som er den klart største kostnaden i hele oppsettet, betaler for de resterende 1 % av drawdown-forbedringen og 0,65 prosentpoeng årlig avkastning, hvorav en del igjen er filtereffekt.

Verdt å merke seg til slutt er at Test A er den eneste av de to som slår referansen i noen periode etter 2009. Den strategien som taper klart samlet, er altså den som har vært best i det siste. Det er et argument for å være varsom med å lese en rangering mellom to strategier som et varig forhold.

Sammenstillingen mot ETF-strategiene i kapittel 4.4 og 4.5 gjenstår. De to funnene der, at V1 slår den enkle indeksswitchen på de fleste mål samtidig som parameterkonfigurasjonen er ustabil ut av utvalget, står fortsatt, og hva de samlet innebærer for strategivalget hører hjemme her og i kapittel 5. Merk at turnover-tallene for ETF-strategiene er oppgitt ensidig og må dobles før de sammenlignes med tallene over, jf. kapittel 3.3.

Spørsmålet om lav volatilitet er en egenskap ved konstruksjonen eller ved parametervalget er besvart i robusthetssveipet sist i kapittel 4.7. Den er en egenskap ved konstruksjonen, siden 29 av 30 kombinasjoner har lavere volatilitet enn ren momentum. Hvilken del av konstruksjonen er besvart i dekomponeringen over, og svaret er sektorbåndet. Samme sveip viser samtidig at rutenettet ikke er robust på drawdown, og at korrelasjonen mellom første og andre halvdel er minus 0,72. Faktorregresjonen i kapittel 3.3 gjenstår, og vil kunne vise om det som er igjen av avkastning er eksponering mot kjente faktorer.

5. Diskusjon og syntese

Fem strategifamilier er testet, altså intraday RSI mean-reversion, filtrerte varianter av den, faktor-ETF-rotasjon på europeiske markeder, absolutt momentum som trendfølging, og to enkeltaksjestrategier på amerikansk marked. Ingen av dem leverte risikojustert avkastning som overgår en passiv referanseportefølje etter kostnader og skatt. Det er svaret på det utvidede forskningsspørsmålet i kapittel 1.2, og det er entydig nok til at det bør sies før nyansene.

5.1 Hva de fem familiene viste

Hypotese 1 er falsifisert. Baseline-varianten av intraday RSI ga minus 13,13 % over perioden, med 85 % av posisjonene lukket på time-stop, altså uten at den forventede bevegelsen materialiserte seg i det

hele tatt. Filtrering løftet resultatet til minus 1,30 %, men fortegnet snudde ikke, og to av de fem filtrene var aktivt skadelige. Falsifiseringen gjelder de testede betingelsene, ikke ideen om mean-reversion som sådan.

Faktorrotasjonen var den første strategien som besto en formell robusthetstest, og den er samtidig den beste illustrasjonen av hvor lite det betyr. Alle tolv parameterkombinasjoner ga positiv Sharpe, spredningen var lav, og ingen kombinasjon hadde dramatisk drawdown. Men edgen satt i drawdown og ikke i avkastning: 12 av 12 varianter hadde lavere fall enn kjøp-og-hold, mens bare 4 av 12 slo den samme referansen på Sharpe, og gjennomsnittlig avkastning var lavere. Og korrelasjonen mellom første og andre halvdel av perioden var 0,25.

Absolutt momentum ble testet som den enklest mulige baselinen, for å se om kompleksiteten i rotasjonen tilførte noe. Den gjorde det. Rotasjonen slo den enkle indeksswitchen på avkastning og på Sharpe. Men begge tapte mot kjøp-og-hold, og trendfølgingens svakhet kom tydelig fram i 2020, der strategien gikk til kontanter i februar og ikke tilbake før i november, altså etter at det meste av opphenting var over.

De to enkeltaksjestrategiene er de mest omfattende testene i rapporten, og de er de eneste som er kjørt over 27 år med punkt-i-tid-data. Ren momentum ga 4,32 % årlig avkastning mot referansens 8,46 %, med en drawdown på 86 %. Multifaktor ga 6,40 % med en drawdown på 58 %. Multifaktor slår ren momentum på hvert måltall. Ingen av dem slår referansen på noe måltall.

5.2 Mønsteret på tvers

Tre trekk går igjen i alle fem familiene, og de er mer interessante enn enkeltresultatene.

Det første er at der en strategi hadde noe å tilby, satt det i risikoreduksjon og ikke i meravkastning. Faktorrotasjonen reduserte fall og ga fra seg avkastning. Multifaktor tjente ikke mer enn ren momentum. De aritmetiske gjennomsnittene er identiske, 8,06 mot 8,04 % ved 100 000 NOK, og hele forskjellen i sluttverdi kommer av at volatiliteten er 18 % mot 26 %. Det er en ekte og verdifull egenskap, men det er en annen egenskap enn den hypotesen handlet om.

Og risikoreduksjonen kommer ikke derfra man skulle tro. Dekomponeringen i kapittel 4.8 skiller sektornøytraliserings fra de fundamentale blokkene, og sektorbandet alene står for 99 % av forbedringen i drawdown og 79 % av forbedringen i volatilitet. Av forspranget på 2,08 prosentpoeng i årlig avkastning kommer 0,79 fra sektorbandet, 0,63 fra kvalifiseringsfilteret og 0,65 fra selve multifaktorscoren. Det som virker, er altså i hovedsak at porteføljen spres over bransjer. En konsentrert momentumportefølje på femten navn havner i én bransje om gangen, og det er dét som produserer fallene på 86 prosent.

Det andre er at parametervalg ikke lot seg overføre i tid. Faktorrotasjonen hadde en korrelasjon på pluss 0,25 mellom halvdelene, og det ble den gang kalt det sterkeste enkeltbeviset mot strategien. Multifaktor har minus 0,72. Det er ikke fravær av sammenheng, det er en invers sammenheng. De parametrene som virket best i første halvdel virket systematisk dårligst i andre. I begge tilfellene ville

et valg tatt på halve historikken pekt feil vei. Det er det sterkeste enkeltargumentet i hele rapporten mot å velge en konfigurasjon på historiske data.

Det tredje er at kostnads- og skattemodellen er en større faktor enn strategivalget ved de kapitalstørrelsene HULC AS opererer med. Friksjonen koster ren momentum 2,5 prosentpoeng i året og multifaktor 2,1, altså rundt en fjerdedel av bruttoavkastningen for begge. Kommisjongulvet på én dollar per ordre binder ved disse ordrestørrelsene, og utgjør dobbelt så stor andel av omsetningen ved 50 000 NOK som ved 100 000. En strategi som ser lønnsom ut brutto kan altså være ulønnsom netto utelukkende på grunn av porteføljestørrelsen, uten at noe ved strategien er endret.

5.3 Hva målingen selv lærte oss

Det mest overførbare i dette arbeidet er ikke resultatene, men hvor mange av dem som hang på infrastrukturen framfor på strategien.

Fire ganger var en test grønn på en regel den ikke utøvde, som beskrevet i kapittel 3.4, og ingen av dem ville blitt funnet ved kodegjennomgang. En femte feil, at porteføljemotoren hoppet over rebalanseringsdatoer som falt i helg, ble ikke funnet av noen test i det hele tatt. Den ble funnet fordi radtallet i en resultatfil ikke stemte med antall datoer i teksten. Den feilen alene endret ren momentum med 1,16 prosentpoeng i året og snudde fortegnet på en konklusjon om rangeringsbåndet.

Datagrunnlaget ville gitt et annet og bedre svar hvis det var valgt billigere. Målingen i kapittel 3.2.1 viser at en gratis kilde mangler 45 % av indeksmedlemmene fra 2005, at nesten alle de manglende er avnoterte, og at kilden viser den korrigerte regnskapsversjonen i to av tre tilfeller der en korleksjon finnes. Begge feilene trekker samme vei, og begge ville sett ut som en bedre strategi. Det er verdt å merke seg at ingen av dem gir en feilmelding.

Og skjevheter som ikke lar seg fjerne, lar seg i det minste måles og retningsbestemmes. Survivorship gjennom manglende datadekning favoriserer multifaktor relativt til ren momentum, altså den strategien sammenligningen handler om, med en relativ risiko på 2,73 og p lik 2,1 ganger 10 opphøyd i minus 7. Størrelsesordenen er samtidig liten nok til at effekten på avkastningstallene sannsynligvis er marginal. Begge deler hører med.

5.4 Hva rapporten ikke kan si

Fraværet av statistisk signifikans er ikke bevis på at strategiene er like. Meravkastningen til multifaktor over ren momentum har p lik 0,96, og multifaktor mot referansen har p lik 0,50. Med et månedlig standardavvik på 3 % i differansen kreves langt flere enn 326 observasjoner for å påvise en forskjell på et par prosentpoeng. 27 år er mye i praksis og lite statistisk, og utvalget kan ikke avgjøre spørsmålet i noen retning.

Hele volatilitetsargumentet hviler dessuten på tre stressepisoder: 2000, 2008 og 2020. Alle drawdown-tallene i rapporten er i realiteten tre observasjoner. Multifaktor sitt fortrinn ble i praksis vunnet i én av dem. Uten dot-com-episoden er det ikke mye igjen av forskjellen mot ren momentum,

og et argument som hviler på én hendelse kan ikke styrkes ved å teste flere parametre på den samme hendelsen.

Faktorregresjonen som er beskrevet i kapittel 3.3 er ikke kjørt. Den ville kunne vise om det som er igjen av avkastning i multifaktor er eksponering mot kjente faktorer, som kan oppnås billigere gjennom indeksfond. Den kan ikke endre hovedresultatet, siden strategien allerede taper mot referansen, men den mangler.

Det er verdt å være eksplisitt om hvilken retning datavalget trekker. Målingene i kapittel 3.2.1 viser at billigere kilder gir for høy avkastning og for gode faktorverdier, altså at svakere datagrunnlag systematisk får strategier til å se bedre ut enn de er. Hovedkonklusjonen i denne rapporten er at ingen av strategiene slår referansen. En eventuell gjenværende datasvakhet ville dermed pekt i retning av at resultatene her er for gode, ikke for dårlige, og kan derfor ikke forklare konklusjonen bort. Det stedet datakvalitet faktisk kan endre et funn, er i sammenligningen mellom de to enkeltaksjestrategiene, og den skjevheten er målt og retningsbestemt i kapittel 3.2.2.

Til slutt gjelder de generelle forbeholdene i kapittel 3.5 fortsatt. En backtest reflekterer ikke latency, delvis fylte ordrer, markedspåvirkning eller disiplin under press, og at en spesifikk implementering ikke har edge er ikke det samme som at tilnærmingen er feil.

5.5 Konsekvenser for de øvrige områdene

Arbeidet med de systematiske strategiene gir tre ting videre til de øvrige områdene.

Punkt-i-tid-kontrakten og porteføljemotoren er infrastruktur som ikke er bundet til de strategiene som er testet her. Enhver strategi som rangerer selskaper på regnskapstall, altså kjernen i fundamental analyse, kan kjøres gjennom det samme laget og arver dermed no-look-ahead, historisk indeksmedlemskap og eksplisitt kostnads- og skattemodellering. Det er den varige leveransen fra dette området.

Kostnads- og skattetallene er direkte overførbare til risikostyring og porteføljekonstruksjon. At friksjonen tar rundt en fjerdedel av bruttoavkastningen ved disse kapitalstørrelsene, og at skatt er den største enkeltmekanismen ved lav turnover, er en beregningsforutsetning for enhver porteføljekonstruksjon i selskapet, ikke et resultat som gjelder bare her.

Og metodikken for å vurdere et resultat er overførbar. Et forhåndsdefinert kriteriesett, et parametersveip framfor et enkeltpunkt, en deling i første og andre halvdel, og en eksplisitt vurdering av om et funn er stort nok til å skilles fra støy. De fire stegene avdekket i denne rapporten flere svakheter enn selve avkastningstallene gjorde.

6. Konklusjon og veien videre

Forskningsspørsmålet var om det finnes algoritmiske handelsstrategier som, gitt realistiske transaksjonskostnader og skatteforhold for et norsk investeringsselskap, kan levere risikojustert

avkastning som overgår en passiv referanseportefølje. Svaret er nei for de strategiene som er testet her. Ingen av de fem familiene slo sin referanse på risikojustert avkastning etter kostnader og skatt.

Hypotese 1, at filtrerte RSI-baserte mean-reversion-strategier kan gi positiv risikojustert avkastning etter transaksjonskostnader, er falsifisert for de testede betingelsene. Utvidelsen til fire andre strategifamilier endret ikke konklusjonen, men gjorde den bredere underbygget.

6.1 Hva som står som funn

Fire ting står igjen som funn framfor som valg, altså som resultater som holder på tvers av parametervalg og ikke bare i den ene kjøringen de ble observert i.

Sektornøytralisering senker risikoen kraftig, og fundamentale blokker legger lite til utover det. Rangering innenfor bransjegruppe står for 99 % av forbedringen i drawdown og 79 % av forbedringen i volatilitet mellom ren momentum og multifaktor. Tilfeldige grupper reproducerer bare 19 og 30 % av de to, så det er den faktiske bransjeinndelingen som gjør arbeidet. At 29 av 30 kombinasjoner i parameterrutenettet har lavere volatilitet enn ren momentum, er en egenskap ved konstruksjonen, men det er sektorleddet i konstruksjonen som bærer den.

Flere navn er bedre enn færre, både på avkastning og på fall. Konsentrasjon på femten navn koster, og den koster mest i volatilitet.

Momentumblokka bidrar minst i dette universet over hele perioden. Men den var den beste blokka i den siste halvdel, så påstanden gjelder 1999 til 2026 som helhet og ikke framtiden.

Og utvalg av parametre på historiske data peker feil vei. Korrelasjonen mellom første og andre halvdel er minus 0,72 for multifaktor og pluss 0,25 for faktorrotasjonen. I begge tilfeller ville et valg tatt midtveis gitt en dårligere konfigurasjon enn et tilfeldig valg.

6.2 Anbefaling

Ingen enkeltkombinasjon forfremmes til produksjon på dette grunnlaget. Kvalitet pluss verdi med 25 navn er den eneste konfigurasjonen i rapporten som slår referansen, med 9,27 % mot 8,46 % og en grunnere drawdown, men den er valgt etter at alle tretti kombinasjoner var kjørt, i et rutenett der utvalg på historiske data beviselig peker feil vei. Å plukke den ville vært å gjøre nøyaktig det rapporten dokumenterer at man ikke skal gjøre.

Den passive delen av porteføljen beholdes derfor som den er, i EØS-hjemmehørende UCITS-fond under fritaksmetoden. Det er ikke en resignasjon. Referansen har i dette utvalget både høyest aritmetisk avkastning og lavest volatilitet, altså begge deler, og en aktiv strategi må slå den på begge for å rettferdiggjøre kompleksiteten og risikoen ved å drifte den.

Dekomponeringen peker likevel på ett grep som er billig nok til å vurderes for seg. Sektornøytralisering krever ingen fundamentaldata, bare en bransjeklassifisering utledet av en offentlig innrapportert kode, og den leverer nesten hele fallreduksjonen. Skulle selskapet på et senere

tidspunkt kjøre en systematisk aksjestrategi, er spredning over bransjer det første som bør ligge inne, og det bør ligge inne uavhengig av om det finnes et fundamentaldatasett i bunnen. Det motsatte gjelder også. Et kommersielt datasett er den klart største kostnaden i oppsettet, og det betaler i dette utvalget for 0,65 prosentpoeng i året, hvorav en del igjen er filtereffekt og ikke signal.

6.3 Veien videre

Neste steg bør ikke være flere varianter på de samme 27 årene. Hele volatilitetsargumentet hviler på tre stressepisoder, og multifaktor sitt fortrinn ble vunnet i én av dem. Flere parametre testet på den samme hendelsen gir ikke ny informasjon.

Det bør i stedet være et uavhengig utvalg, altså et annet univers. For HULC AS er det dessuten det relevante universet, nemlig europeiske noteringer under fritaksmetoden. Et multifaktorsveip på et EØS-univers ville gitt både en ny stikkprøve og et resultat som faktisk kan handles. Holder mønstrene fra kapittel 4.7 der, er de verdt å tro på. Holder de ikke, er dette rutenettet forklart som et regimeutslag.

Dekomponeringen endrer også hva som bør testes først på det nye universet. Det mest informative er ikke en ny multifaktorvariant, men å kjøre den samme trappen der, altså ren momentum, momentum med sektorband, og multifaktor. Holder fordelingen mellom leddene seg, er sektorspredning et generelt funn og fundamentaldataene et dyrt tillegg. Snur den, er dette utvalget forklart.

Tre målinger gjenstår før bildet er komplett. Faktorregresjonen i kapittel 3.3 er ikke kjørt, og bør kjøres på begge enkeltaksjestrategiene. Den europeiske kommisjonsulempen er ikke målt, og markedssvalget i kapittel 3.1.3 står som en åpen metodisk post inntil den er det. Og terskelen på 300 000 NOK for å revurdere markedssvalget bør erstattes av en beregnet terskel når begge effektene kan måles i samme simulering.

Det som uansett følger med videre, er rammeverket. Punkt-i-tid-kontrakten, porteføljemotoren med kostnads- og skattemodellering, og valideringsmetodikken med mutasjonstesting og forhåndsdefinerte kriterier er bygget for å kunne bære andre strategier enn de som er testet her. Rapportens tydeligste lærdom er at et backtestresultat i stor grad er en egenskap ved måleapparatet, og at et apparat man kjenner svakhetene til er mer verdt enn et resultat man ikke kan etterprøve.

7. Data og reproduserbarhet

Enkeltaksjestrategiene er kjørt på S&P 500-medlemmer med historisk indeksmedlemskap, 55 rebalanseringer fra 30. juni 1999 til 30. juni 2026, med fundamentaldata i opprinnelig rapportert form filtrert på tilgjengelighetsdato. ETF-strategiene er kjørt på perioden januar 2018 til desember 2024, og intradagstrategiene på januar til juni 2024. Kostnadsmodellen er gjengitt i kapittel 3.1.2 og skattebehandlingen i kapittel 3.1.3. Alle parametervalg er oppgitt i den seksjonen som presenterer den enkelte testen. Rammeverket er versjonskontrollert, og hver måling er kjørt mot en full testsuite.

Egne måleresultater er ikke referert til kilder, siden de er produsert i dette arbeidet. Kildelisten nedenfor dekker de påstandene rapporten låner fra andre, altså teori, tidligere empiriske funn, definisjoner av nøkkeltall, kommisjonssatser og lovhjemler.

8. Referanser

Artikler i tidsskrift

- Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24(1), 34-112.
- Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3), 929-985.
- Bailey, D. H., Borwein, J. M., López de Prado, M., & Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 61(5), 458-471.
- Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4), 553-580.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.
- Da, Z., Liu, Q., & Schaumburg, E. (2014). A closer look at the short-term return reversal. *Management Science*, 60(3), 658-674.
- Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221-247.
- De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3), 793-805.
- Faber, M. T. (2007). A quantitative approach to tactical asset allocation. *The Journal of Wealth Management*, 9(4), 69-79.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2), 153-193.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.
- Fama, E. F., & French, K. R. (2021). The value premium. *The Review of Asset Pricing Studies*, 11(1), 105-121.
- Figuerola-Ferretti, I., Bermejo Climent, R., Santos Moreno, Á., & Hevia, T. (2021). Factor investing: A stock selection methodology for the European equity market. *Heliyon*, 7(10).
- Harvey, C. R., & Liu, Y. (2015). Backtesting. *The Journal of Portfolio Management*, 42(1), 13-28.
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5-68.

- Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3), 881-898.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91.
- Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4), 917-926.
- Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1), 1-28.
- Ljungqvist, A., Malloy, C., & Marston, F. (2009). Rewriting history. *The Journal of Finance*, 64(4), 1935-1960.
- McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1), 5-32.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1), 1-28.
- Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 21(1), 49-58.
- Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3), 59-64.

Bøker

- Antonacci, G. (2014). *Dual momentum investing: An innovative strategy for higher returns with lower risk*. McGraw-Hill.
- Connors, L., & Alvarez, C. (2009). *Short term trading strategies that work*. TradingMarkets Publishing Group.
- Wilder, J. W. (1978). *New concepts in technical trading systems*. Trend Research.

Data og programvare

- Aroussi, R. (u.å.). *yfinance* [Python-bibliotek]. <https://github.com/ranaroussi/yfinance>
- French, K. R. (u.å.). *Detail for 12 industry portfolios*. Data Library, Tuck School of Business, Dartmouth College. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
- Interactive Brokers. (u.å.). *Commissions - stocks*. <https://www.interactivebrokers.com/en/pricing/commissions-stocks.php>
- Sharadar. (u.å.). *Core US Equities Bundle* (tabellene SF1, SP500, TICKERS og SEP) [dataset]. Nasdaq Data Link. <https://data.nasdaq.com/databases/SFA>

Lov og regelverk

- Lov om skatt av formue og inntekt (skatteloven), LOV-1999-03-26-14, § 2-38. <https://lovdata.no/lov/1999-03-26-14>