Algorithmic trading strategies from a retail perspective

An empirical assessment of RSI mean reversion, factor rotation and multifactor strategies

HULC AS

Systematic trading strategies - interim report

Noah Lekhal Husnes

Student at BI Norwegian Business School

August 2026

1. Introduction

1.1 Background and motivation

The first book I read about equity trading presented the Relative Strength Index as a tool for identifying oversold stocks ripe for a bounce. The idea is plausible on its face. A mechanical indicator points out buying opportunities, without discretion and without company analysis. It raised a question that this report is an attempt to answer: whether short-term price deviations are measurable and can be exploited systematically.

HULC AS was incorporated in January 2026 as an investment company with the purpose of developing structured investment strategies grounded in financial theory. The company's activity is organised around three complementary areas: systematic trading, fundamental analysis, and risk management and portfolio construction. This interim report deals with systematic trading.

The theoretical assumption underlying the entire project is that the market is predominantly efficient over the long run, but that short-term deviations and mispricings arise regularly and can potentially be exploited systematically. This interim report attempts to test the latter part of that assumption empirically.

1.2 Research question and hypothesis

The hypothesis was formulated as follows at the outset of the work:

> *Hypothesis 1: Filtered RSI-based mean reversion strategies can deliver positive risk-adjusted returns after transaction costs.*

This hypothesis is the point of departure for the work. When empirical results subsequently showed that the hypothesis could not be confirmed, the investigation was widened to cover alternative algorithmic approaches. That extension is not a departure from the original plan, but a logical consequence of the research method. When a hypothesis is falsified, research turns to alternative explanations or related hypotheses that can be tested.

The extended research question can be stated as follows:

> *Are there algorithmic trading strategies that, given realistic transaction costs and tax conditions for a Norwegian investment company, can deliver risk-adjusted returns exceeding those of a passive benchmark portfolio?*

The capital levels used throughout the report, NOK 50,000 and NOK 100,000, are not a simplification. They are part of the question. The literature that motivates these strategies is written largely on institutional assumptions, where the order is large enough that fixed per-order costs are immaterial. That assumption does not hold for a small company, and it holds even less for a private individual. Both trade in the same market as participants whose capital, infrastructure and analytical resources are of an entirely different order of magnitude. The report therefore tests not only whether the strategies have an edge, but whether the edge survives the cost of being small. That is a different question from the one the literature asks, and the results in Chapter 4 give a different answer.

1.3 Structure of the report

The report is structured as follows. Chapter 2 presents the theoretical framework and a review of the strategies identified as relevant to test, based on a combination of traditional financial theory, academic research and pragmatic considerations relating to the practical constraints facing a retail investor. Chapter 3 describes the methodology applied, including the backtesting framework, the cost modelling and the statistical measures used. Chapters 4 to 6 present results, interpretation and conclusions, together with a discussion of the implications for the two remaining areas.

1.4 Scope and delimitations

The report covers only systematic, rule-based strategies. Strategies that require fundamental analysis of individual companies, or overall portfolio construction and risk management, belong to the two other areas and are not part of this interim report. They will be addressed in separate documents.

The report is further limited to strategies that are practically feasible for HULC AS as a retail participant. This means that strategies presupposing institutional infrastructure, low-latency access or market data beyond what is available through Interactive Brokers have not been considered.

It is also important to emphasise a fundamental resource constraint. As a small Norwegian investment company, HULC AS has limited access to expensive data sources, professional infrastructure and dedicated development time. The report therefore covers what we judge to be the most representative and relevant broad approaches within systematic trading - mean reversion, momentum, trend following, factor rotation and multifactor strategies - but does not treat each approach exhaustively. This delimitation affects how categorical our conclusions can be.

2. Theoretical framework and strategy selection

2.1 The efficient market hypothesis

A natural starting point for any assessment of algorithmic trading strategies is Eugene Fama's efficient market hypothesis (Fama, 1970). The hypothesis holds that financial markets reflect all available information in their prices, and that it should therefore not be possible to obtain systematic excess returns through analysis of historical prices or publicly available information. Fama presented three forms of the hypothesis: weak, semi-strong and strong efficiency.

Algorithmic trading strategies based on technical indicators such as the RSI are a direct test of weak-form market efficiency. If weak-form efficiency holds, historical price patterns cannot be used to predict future price movements, and strategies based on such patterns should consequently not deliver systematic excess returns after costs.

There is, however, extensive empirical research documenting deviations from market efficiency, particularly over short horizons. These deviations are commonly referred to as anomalies in the academic literature, and include the momentum effect, short-term mean reversion, and various factor premia such as value, quality and low volatility. The anomalies are the reason systematic strategies are still researched and traded despite the theoretical expectation that they should not work.

2.2 Mean reversion as a theoretical framework

Mean reversion is the hypothesis that asset prices tend to revert towards a statistical average over time. The hypothesis was systematically explored by, among others, Jegadeesh (1990) and Lehmann (1990), who documented that stocks which had underperformed over short horizons (typically one week to one month) exhibited statistically significant reversal in the following period. De Bondt and Thaler (1985) demonstrated corresponding effects over longer horizons.

The RSI, developed by J. Welles Wilder (1978), is one of the most widely used technical indicators for operationalising mean reversion. The indicator measures the strength of recent price movements and normalises it to a scale from 0 to 100. By convention, values below 30 are taken to indicate an oversold condition and values above 70 an overbought condition. A simple mean reversion strategy consists of buying when the RSI falls below 30, in the expectation that the price will normalise towards its average.

Despite its theoretical appeal and intuitive logic, more recent research shows that the edge associated with pure RSI mean reversion has weakened substantially in modern markets. This is attributed to

several factors: broader institutional participation, higher market efficiency, and in particular the fact that the strategy has become so well known that it affects price formation in precisely the cases where it generates signals. The mechanism is widely documented. McLean and Pontiff (2016) find that the returns to published anomalies fall substantially after publication.

2.3 Strategies considered for testing

Based on the theoretical review and practical considerations, the following strategies were identified as relevant to test. They represent varying degrees of complexity and are motivated by different aspects of the overarching hypothesis. Together they cover the main categories of systematic trading strategies available to a retail participant: mean reversion, momentum (both time-series and cross-sectional), trend following and multifactor approaches.

2.3.1 Intraday RSI mean reversion

The original strategy from the project proposal, based directly on Hypothesis 1. The strategy identifies stocks that have fallen sharply at the NYSE open and that have an RSI below 30 on intraday bars. Positions are opened with fixed percentage-based bracket orders (3% take profit, 1.5% stop loss) and closed by a time stop after 90 minutes if neither bracket is triggered.

The strategy represents the simplest implementation of the hypothesis and serves as the baseline for further work. It is motivated by traditional technical analysis and is a direct operationalisation of the RSI literature.

2.3.2 Filtered intraday variants

When the baseline strategy showed no edge, five filters were identified in the literature as potential improvements. These were tested both in combination and individually, in order to isolate the effect of each filter:

- Trend filter: take an entry only when the share price is above its 200-day moving average. Motivated by research indicating that mean reversion works best in stocks that are in an established uptrend (Connors and Alvarez, 2009).
- Gap size filter: take an entry only when the opening decline falls within a given interval (-8% to -3%). Motivated by the observation that moderate declines are often flow-driven whereas extreme declines are often fundamentally driven (Da, Liu and Schaumburg, 2014).
- Reversal confirmation: wait until a green 5-minute bar has been observed before entering. Motivated by the fact that one then buys an actual reversal rather than the continuation of a downtrend.
- ATR-based SL/TP: replace fixed percentage stops with stops based on the Average True Range. Motivated by the fact that volatility-adjusted stops give more consistent risk exposure across stocks.
- Risk-based position sizing: calculate the number of shares such that the dollar risk per trade is a fixed percentage of the portfolio. Motivated by the Kelly criterion (Kelly, 1956) and modern risk management.

2.3.3 Cross-sectional momentum in single stocks

A classic and well-documented strategy in which stocks are ranked by their return over a given period (typically 12 months excluding the most recent month), and the portfolio holds the highest-ranked names. The strategy was first documented by Jegadeesh and Titman (1993) and has been the subject of extensive subsequent research.

The strategy was initially assessed as not applicable to HULC AS. The reasoning was that US equities fall outside the Norwegian participation exemption, so that frequent realisation of gains carries a structural cost of 22% of the realised gain. That assessment led to single-stock strategies being moved to EEA instruments.

This assessment has since been revised. It weighed the tax cost in isolation, without setting it against the other structural costs at the capital level the strategy would actually be run with. The calculation is worked through in Chapter 3.1.3. The conclusion there is that the ranking between markets cannot be settled on the evidence available. The tax effect has been measured; the European commission disadvantage has not, and the choice of market stands as an open methodological item. Cross-sectional momentum in US single stocks has therefore been brought back into the strategy set, and the justification for running the tests on the US market now rests mainly on data availability, as discussed in Chapter 3.2.1. At the same time, the strategy serves as a control group for the multifactor strategy in Chapter 2.3.6, in that it shows what momentum alone delivers on the same universe.

2.3.4 Factor ETF rotation for European markets

Inspired by Asness, Moskowitz and Pedersen (2013), who documented that factor premia such as value and momentum exist across asset classes and markets, and by more recent research on factor momentum, a rotation strategy across European factor ETFs was identified as a pragmatic and tax-efficient approach.

The strategy uses UCITS-domiciled factor ETFs that fall within the Norwegian participation exemption for a Norwegian investment company. Three variants were formulated and tested:

- Variant 1 - Pure relative momentum: monthly rebalancing, holding equal weights in the two ETFs with the highest 6-month return.
- Variant 2 - Faber-style (absolute + relative momentum): as Variant 1, but only ETFs trading above their 200-day SMA qualify. If fewer than two qualify, the remainder is held in a bond ETF.
- Variant 3 - Volatility-weighted with a momentum tilt: the weighting of qualifying ETFs is inversely proportional to their volatility, adjusted for momentum ranking.

2.3.5 Absolute momentum (trend following)

A simpler form of trend following, classically described by Faber (2007) and developed further by Antonacci (2014). The strategy holds a broad market portfolio (represented by the IMEU ETF) as long as it trades above its long-term moving average, and switches to a defensive allocation (government

bonds) when the market breaks below it. The strategy is motivated by the observation that most large market losses occur in prolonged downtrends that are identifiable in advance.

The strategy serves several purposes in this study: it acts as the simplest possible benchmark for the trend-following principle, and its result can be used to assess whether the complexity of the factor rotation strategies adds anything beyond a simple switch between equities and bonds.

2.3.6 Multifactor strategy (Quality-Value-Momentum) in single stocks

Research on European markets, documented among others by Figuerola-Ferretti, Bermejo Climent, Santos Moreno and Hevia (2021), covering the period 1991-2019 for European large-cap equities, indicates that multifactor approaches combining quality, value and momentum in single stocks have historically delivered stronger risk-adjusted returns than single-factor strategies.

The strategy operates on the following principle: each stock in the universe is given a standardised score on each of the three factors, the scores are combined into a single composite ranking, and the portfolio holds the highest-ranked companies at equal weight. Quality is measured by profitability and leverage, where gross profit over assets follows Novy-Marx (2013); value by pricing multiples; and momentum by the return over 12 months excluding the most recent month. Standardisation is performed within sector, so that the ranking does not in practice become a sector allocation. The portfolio is rebalanced semi-annually.

Testing QVM places stricter demands on the underlying data than the other strategies in this report. The requirements are point-in-time fundamental data and an equity universe free of survivorship bias. These requirements can be met for US equities within the constraints of the project, but not for European ones. Together with the cost assessment in Chapter 3.1.3, this is the reason the strategy is tested on the US market. The data requirements are discussed in Chapter 3.2.1.

2.4 Rationale for the strategy selection and its limitations

The strategies selected for testing are not a random sample. They represent a logical progression: from the original hypothesis (RSI mean reversion), through incremental refinements (filtered variants), to structurally different approaches (cross-sectional momentum, factor rotation, multifactor in single stocks). This progression reflects how empirical research typically develops: a hypothesis is tested, and if it does not hold, one either seeks to refine it or explores related hypotheses.

Taken together, the strategies cover the five main categories within systematic trading that are practically feasible for a retail participant: mean reversion (intraday RSI), cross-sectional momentum (single stocks), time-series momentum (factor rotation), trend following (absolute momentum), and multifactor (QVM). This gives the report breadth of coverage, even though no individual category is tested exhaustively.

Several strategies have not been tested in full in this report for resource reasons:

- Statistical arbitrage and pairs trading: requires sophisticated statistical analysis and infrastructure for continuous monitoring of a large number of stock pairs.

- Currency carry trades: requires tolerance for extreme volatility and is not immediately relevant to the profile of HULC AS.
- Volatility strategies: selling volatility through options and related instruments requires specialist expertise and an options account.
- Seasonal strategies: the effect is typically so small that they work best as filters layered on top of other strategies.
- More exotic momentum variants such as residual momentum, factor momentum and industry momentum.

It has also been a deliberate choice not to test every conceivable parameter variant for each strategy. Such an approach would entail a considerable risk of data mining, whereby a sufficient number of attempts will always produce some variants that appear profitable by chance alone (Harvey, Liu and Zhu, 2016; Bailey et al., 2014). Instead, each variant is motivated by specific theoretical or practical reasoning before it is tested, and parameter robustness is examined systematically for the most promising strategy.

The combined consequence of these delimitations is that the report can draw conclusions about the strategies tested under the conditions tested, but not about the class of algorithmic strategies in general. We have touched on all the broad approaches, but as of today we do not have the resources to treat each of them exhaustively. On this basis the ground is laid for the conclusion presented in Chapter 6, where we summarise what the results actually tell us and what the rational way forward is for HULC AS given that knowledge.

3. Methodology

3.1 The backtesting framework

A dedicated backtesting framework has been developed in Python in order to test the relevant strategies under consistent conditions. The framework is structured modularly, with separate components for data loading, strategy implementation, order execution simulation, portfolio management and report generation. All code is under version control, and every component is covered by unit tests that are run in full before each measurement.

The framework enforces several fundamental principles that are necessary for backtest results to be credible:

3.1.1 Enforcing no look-ahead

One of the most common problems in amateur backtests is so-called look-ahead bias, where the strategy inadvertently gains access to information that would not have been available at the time of the trade. In this framework the constraint is enforced technically. The strategy can only see data from the period strictly before the trading day in question. A specific test has been written that attempts to breach this principle by means of a spy strategy, and that verifies that the framework does in fact catch the attempt.

Price data, however, is only one of three sources of look-ahead for strategies operating on single stocks. The other two are restatements, that is, accounting figures being corrected later and the corrected version being read back in time, and publication lag, that is, an accounting figure being made available to the strategy from period end rather than from the actual publication date. A third, related source of error is universe construction, where companies that were subsequently delisted are omitted from the periods in which they were in fact index members. Ljungqvist, Malloy and Marston (2009) document how historical databases are rewritten after the fact without this being visible to the user.

The framework has therefore been extended with a dedicated contract governing access to fundamental data and universe lookups. The contract requires that every request specify a decision date, that only observations available on or before that date be returned, that only originally reported values be visible, and that the universe be built on historical index membership. Missing data produces an explicit error rather than an empty result, so that absence cannot be mistaken for a valid observation. The contract is vendor-agnostic, so that vendor-specific logic is confined to an adapter layer and kept out of the strategies.

What the contract does in practice can be illustrated with a single company. Lehman Brothers was a member of the S&P 500 up to and including the 30 June 2008 snapshot, and filed for bankruptcy on 15 September that year. On the decision date of 30 June 2008 the contract returned a return on equity of 15.4%, net income of USD 3,461 million and shareholders' equity of USD 24,832 million. The two preceding years gave 20.5% and 21.2%. Ten weeks after the last observation the company no longer existed.

A quality ranking on that day should have placed Lehman among the better names in the universe. That is the correct answer, because that was the information available. A backtest that had instead used the later corrected figures, or that had excluded Lehman from the universe because the company does not exist today, would have reported a return no investor could have achieved.

The example illustrates the publication lag just as concretely. The figures available on 30 June 2008 related to the accounting period ending 29 February 2008 and became available on 9 April. They were 82 days old on the decision date. Filtering on period end rather than availability date would have given the strategy figures it could not have seen.

3.1.2 Realistic cost modelling

All simulated trades are charged costs reflecting the actual terms at Interactive Brokers Pro (Interactive Brokers, n.d.):

- Commission: USD 0.005 per share, minimum USD 1.00 per order, maximum 1% of trade value.
- Bid-ask spread: 3 basis points per side, modelled as buys executing above and sells below the mid price.
- Slippage: 5 basis points per side, in addition to the spread. This is a conservative estimate based on actual trades in liquid US equities and European ETFs.

Total costs per round-trip trade therefore amount to approximately 16 basis points plus the flat commission on small orders. This is deliberately set somewhat conservatively in order to avoid overly optimistic results. An earlier version of the framework used a more aggressive slippage model with a volatility coefficient, but this was revised when it turned out to introduce disproportionately large costs for the ETF strategies. The significance of the minimum commission at small order sizes is discussed in Chapter 3.1.3.

3.1.3 Tax and structural trading costs

An important methodological consideration is that HULC AS, as a Norwegian limited company, is subject to corporate taxation. For equity investments outside the EEA, an effective tax rate of 22% applies to realised gains. For shares and funds domiciled within the EEA (under the Norwegian participation exemption), capital gains tax is largely waived for corporate investors, cf. Section 2-38 of the Norwegian Taxation Act.

This distinction has material consequences for the choice of strategy and market, and was decisive in the project moving early on from US to European instruments. In isolation, the tax difference means that a gross return of 12% CAGR is reduced to 9.36% CAGR if the entire gain is realised on an ongoing basis outside the participation exemption.

Tax, however, is only one of several structural costs, and it is not necessarily the largest at the capital level at which HULC AS operates. Two circumstances make the picture more complex than the original assessment assumed.

First, the participation exemption does not fully cover a European equity universe. The exemption applies to companies domiciled within the EEA. The United Kingdom is no longer an EEA member, and Switzerland never has been. These two markets make up a substantial share of the Stoxx 600. A broad European single-stock universe therefore carries an effective tax rate on realised gains in the order of 7 to 8%, not zero. The tax advantage relative to the US market is real, but considerably smaller than 22 percentage points.

Second, the commission structure differs between markets. Interactive Brokers applies a minimum charge per order. On US equities the minimum is USD 1.00. On European exchanges the minimum is several times higher, and UK shares additionally attract stamp duty of 0.5% on purchase. The minimum charge is a fixed cost per order and therefore accounts for a higher share the smaller the order is.

The consequence is that the ranking between markets depends on order size. With an allocation of NOK 50,000 to single-stock strategies spread across 10 to 20 positions, each order is in the order of NOK 2,500 to 5,000. At such order sizes the minimum commission accounts for a high share of order value, and a European universe would carry a cost drag that a US one does not.

The original assessment assumed that this cost drag exceeds the tax advantage by a comfortable margin, and that the choice of market therefore followed from the cost picture. That assessment does not stand on its own. The tax effect has now been measured explicitly in the portfolio layer, and it is considerably larger than the entire cost model at the capital levels in question. The relative size of the two moreover

depends on how much the portfolio turns over, and is therefore not a constant that can be used to rank markets independently of strategy.

Nor is the question settled, for two reasons. The European commission disadvantage has not been measured. The minimum charges at Interactive Brokers and the stamp duty on UK shares are known, but no simulation has been run on a European universe with European commission rates, and it is therefore unknown how large the disadvantage becomes expressed in annual return. In addition, a European portfolio would have a different tax position under the participation exemption, with an effective rate in the order of 7 to 8% rather than 22%. The measured tax effect is calculated at 22% and cannot simply be transferred to a European universe.

The single-stock strategies are therefore tested on the US market, with 22% tax on realised gains modelled explicitly in the backtest. The justification for this choice now rests mainly on data availability, as discussed in Chapter 3.2.1, and not on the cost picture. The passive part of the portfolio continues to be held in EEA-domiciled UCITS funds under the participation exemption. The NOK 300,000 threshold previously stated for reconsidering the choice of market is not derived from measurement. It should be replaced by a calculated threshold once the European commission disadvantage has been measured in the same way as the tax effect has been. Until then, the choice of market stands as an open methodological item.

3.2 Data sources and data quality

Different strategies require different data sources. The following sources have been used:

- Intraday data (5-minute bars) for US equities: obtained via the Interactive Brokers API. This is the same data source that would be used in live trading, which ensures consistency between the backtest and potential production use.
- Daily data for US equities and European ETFs: obtained via yfinance (Aroussi, n.d.), an open Python library that uses Yahoo Finance as its underlying source. Auto-adjustment for dividends and splits is enabled.
- Fundamental data and historical index membership for US single stocks: the Sharadar Core US Equities Bundle, distributed through Nasdaq Data Link. The SF1 table provides accounting figures both as originally reported and in later corrected form, with an availability date for each observation, and SP500 provides historical index additions and deletions. Delisted companies are included. The dataset has been acquired and is in use, and the measurements in Chapters 3.2.1 and 3.2.2 were run against it. The requirements for this type of data are discussed in Chapter 3.2.1.

All data is cached locally as Parquet files to ensure reproducibility and reduce load on the data sources. For Interactive Brokers, historical intraday requests must respect rate limits (a maximum of 60 requests per 10 minutes), and a semaphore-based pacing mechanism has been implemented to comply with these constraints.

Several data quality issues have been identified that have consequences for the test period and for the interpretation of results:

- The momentum ETF IEMO has complete data only from 2018; the 2015-2017 period is thin following launch.
- Certain ticker substitutions were necessary because some ETFs do not have sufficient history on the London Stock Exchange. For example, IEMO.L was replaced with IEMO.MI (the Milan listing), and IMEU.L was replaced with IMEU.MI because the London listing is quoted in GBp and therefore produces erroneous price movements.
- The factor rotation strategy is therefore tested over the period 2018-01 to 2024-12, rather than from the ETFs' original launch dates. This is done to ensure that all strategy variants run on the same data.

3.2.1 Data requirements for the single-stock strategies

The two strategies operating on single stocks place stricter demands on the underlying data than the ETF strategies. Both require a universe built on historical index membership, and the multifactor strategy additionally requires historical fundamental data spanning several years. The data requirement is therefore treated separately here, because it determines both whether the strategies can be tested in a way that yields interpretable results, and in which market the tests can be carried out.

Four requirements must be satisfied for the tests to support conclusions. The first two apply to both single-stock strategies; the last two are particularly critical for the multifactor strategy:

- Point-in-time fundamental data. Accounting figures as they were published at the time of the decision, not as they appear today. In two out of three cases, free sources show the most recently corrected version back in time, as measured below. The strategy thereby gains implicit knowledge of corrections that were not publicly known when the decision was made.
- An equity universe free of survivorship bias. The universe must be built on historical index membership, so that companies that subsequently went bankrupt, were acquired or were delisted are included in the periods in which they were in fact members. A universe assembled from today's index members consists by definition of survivors, and produces artificially good results regardless of the quality of the strategy (Brown et al., 1992). This requirement applies to purely price-based strategies as well.
- Sufficient history. The value factor had an unusually weak period in developed markets from roughly 2010 to 2020 (Fama and French, 2021). A test period of five to six years measures a single regime and provides no basis for drawing conclusions about the factor in general.
- Consistent calculation of key figures. The quality and value metrics must be calculated in the same way across companies, and the definitions must be the same in the backtest as in any subsequent live operation. If the definitions differ, the strategy being run is not the one that was tested.

The claim that free sources are inadequate has been tested directly against the commercial dataset. The sample is 120 randomly drawn S&P 500 members per membership date, 311 unique symbols in total. The share of the index members of the day on which yfinance today has nothing whatsoever falls monotonically with the age of the membership.

Membership date	Sample	Missing	Share

30 June 2005	120	54	45.0%
30 June 2015	120	24	20.0%
30 June 2022	120	10	8.3%

Between 92 and 96% of the missing companies are delisted. This is survivorship in its purest form. In practice the source only has the companies that still exist. A backtest from 2005 built on yfinance would be missing almost half the universe, and the missing names are systematically the worst performers. The depth is moreover four years and not five. yfinance returns five annual columns, but the oldest is empty, and an attempt at fiscal year 2021 gave zero comparable observations out of 595 possible. The source therefore cannot be used for fundamental backtesting beyond four years, regardless of how corrections are handled.

Among the companies that are present, the yfinance value deviates from the originally reported figure by more than 0.5% in 57 of 448 cases for fiscal year 2022, that is 12.7%, and in 64 of 537 for 2023, that is 11.9%. Among those that deviate, the median deviation is 3.06%, the 95th percentile is 23.83% and the maximum is 88.3%.

The deviation rate varies sharply between fields, and the distinction is decisive for interpretation. For total debt, 74 of 197 observations deviate, that is 37.6%, but only two of them match the corrected value. That is a definitional difference, not a timing difference. Among other things, the sources treat lease obligations differently. For revenue, 38 of 197 deviate, that is 19.3%, and 16 of them match the corrected value. For assets, equity and net income the definitions are comparable, and the deviation rates are 0.5%, 1.0% and 3.0% respectively.

For the three fields where the definitions are comparable, 9 of 591 observations deviate, that is 1.5%, and all nine match the corrected value. Turned around: among the 42 observations where the commercial dataset has itself recorded a correction, yfinance shows the corrected value in 27 cases, that is 64.3%, the originally reported value in 11 cases, that is 26.2%, and something else in 4 cases. The matches are exact, not approximate. The source delivers precisely the figure the company reported later.

A backtest on yfinance is therefore hit by two errors at once, and both pull in the same direction. The universe is missing the losers, and where a correction exists, the source shows it in two out of three cases. The first error produces returns that are too high because the bankruptcies are gone. The second produces factor values that are too good because the accounting figures are the revised ones. Neither announces itself in the result. Both look like a better strategy.

How much corrections matter has been measured on the dataset's own as-reported dimension, restricted to companies that were S&P 500 members over the period 1998 to 2026. Of 91,740 accounting periods, 1,321 have more than one filing, that is 1.44%. Measured per company, 724 of 1,159 are affected at least once, that is 62.5%. The median time from original filing to correction is 23 days, the 95th percentile is 106 days and the maximum is 452 days.

The contrast between the two proportions is the point. Corrections are rare per accounting period and almost universal over a company's lifetime. A backtest that reads corrected figures therefore does not get a small and evenly distributed error. It gets correct figures for most periods, and systematically favourable figures for precisely those periods where something went wrong.

The size of the corrections points the same way. Of the six cases used by the test suite, four either flip sign or change net income by more than 90%. Tyco International's result for the quarter ending 31 March 2002 was filed on 15 May as plus USD 1,400 million and corrected on 12 June to minus USD 3,113 million, that is 28 days later. A quality factor calculated on corrected figures would have correctly ranked Tyco as one of the weakest companies in the universe in the spring of 2002. That ranking did not exist when the decision had to be made, and a backtest that uses it is measuring the ability to see into the future.

The requirements can, however, be met for US equities. Commercial point-in-time datasets with delisted companies included, history back to the 1990s and a dedicated table of historical index membership are available at a monthly cost well within the project's budget. Corresponding coverage for European equities requires institutional databases that are not available here. This, together with the cost assessment in Chapter 3.1.3, is the reason the single-stock strategies are tested on the US market.

It should be made clear that fundamental data delivered through a broker's market data subscription does not meet this need. Such services provide current key figures for listed companies, not historical observations as they stood at the time of the decision, and they do not include companies that have subsequently been delisted. They are usable for ongoing operations, but not as a basis for a backtest.

Data access alone does not guarantee that the test is correct. Such datasets deliver accounting figures in several versions, both as originally reported and in later corrected form, and each observation is tagged with the date it became available. Point-in-time correctness is achieved only when the query both selects the originally reported version and filters out observations that did not exist at the time of the decision. An error here produces no visible error message, only a result that is too good. The framework's no-look-ahead test has therefore been extended to cover fundamental data access.

Even with correct data, return figures alone do not determine whether a strategy has value. Positive risk-adjusted returns may be due to exposure to known factors rather than to the composition of the strategy. The results are therefore checked against a factor regression, as described in Chapter 3.3.

3.2.2 Survivorship through data coverage

The contract closes off survivorship in universe construction. The membership lists are historical, and companies that went bankrupt or were acquired are included in the periods in which they were members. The multifactor strategy, however, additionally excludes companies for which fundamental data is missing. If missing coverage is associated with a company being on its way under, the bias reappears through a mechanism other than the one the contract covers. This has been measured.

Across 53 rebalancings, with a 12-month window, 211 names were excluded from the strategy because of missing fields. Of these, 24 disappeared from the index within 12 months, that is 11.4%. In the

universe as a whole, 1,102 of 26,497 names disappeared, that is 4.2%. The relative risk is 2.73. A two-sided z-test gives $z = 5.20$ and $p = 2.1 \times 10^{-7}$. The difference is not consistent with chance. The last two of the 55 rebalancing dates are omitted, because they lack a full twelve-month window.

The finding is less serious than the figure alone suggests, because the 24 exclusions are spread across 15 unique companies and across three different mechanisms.

Mechanism	Count	Companies
Foreign domicile	10	UL, SHEL, INCLF, PDG, B
Restructuring	8	BXLT, CPGX, TFCFA, H1, VNT
Genuine distress	6	SBNY, FRCB, FMCC, FLTWQ, LDWIF

Foreign domicile is the largest group and the least concerning. All were removed on 22 July 2002, when S&P took foreign companies out of the index. They lacked US GAAP figures because they were foreign, and they were removed because they were foreign. The correlation is real, but it concerns accounting standards and not the condition of the company. Restructuring covers acquisitions, spin-offs and mergers where accounting reporting was interrupted by the transaction itself, and the direction there is unclear, since acquisitions occur both at a premium and in distressed sales. Genuine distress is the group that constitutes a bias. Companies stop delivering reportable figures when they are on their way under, and therefore drop out of the strategy just before they collapse.

The direction of the bias is known. It favours the multifactor strategy relative to the purely price-based strategy, that is, the strategy the hypothesis is about. The multifactor strategy is not allowed to own the worst outcomes in the distress group, because they are invisible on the decision date. The price-based strategy ranks only on price and does own them. Both FRCB and SBNY are in the price-based universe at 31 December 2022, with 12-1 momentum of minus 37.9% and minus 55.9% respectively. The comparison between the two strategies is therefore not entirely neutral, and the bias runs in favour of the strategy being tested.

No correction has been attempted. Compensating would have required assumptions about what the missing figures would have been, that is, ascribing a fictitious profitability and leverage to companies that stopped reporting. That would replace a known and measured bias with an unknown model error.

The order of magnitude should stand alongside the significance. 24 company exclusions across 53 rebalancings with roughly 495 qualifying names each amounts to just under 0.05% of all position opportunities. In a 15-name portfolio the probability that one of them would actually have been selected is small, and they would in any case have ranked poorly on the quality factor. The bias is statistically certain, but the effect on the return figures is probably marginal. Both facts are part of the finding.

3.2.3 Construction of the multifactor score

The multifactor strategy is the most data-intensive of the tests, and the construction of the score itself involves several choices, each of which can change which companies are selected. They are collected here, because a result cannot be interpreted without them.

The score is built from eight metrics distributed across three blocks. A company must have current values in twelve fundamental fields to be ranked at all, and qualification takes place in the data layer before the strategy sees the universe. A company that drops out ends up on an exclusion list with a reason attached, not in silence.

Metric	Block	Direction
Return on equity, TTM	Quality	Higher is better
Gross profit / assets	Quality	Higher is better
Debt / equity	Quality	Lower is better
Net income / market cap	Value	Higher is better
Book equity / market cap	Value	Higher is better
Operating income / enterprise value	Value	Higher is better
Free cash flow / market cap	Value	Higher is better
Return 12m excl. most recent month	Momentum	Higher is better

All metrics are oriented to 'higher is better' in one place in the code. Scattered sort directions are the most likely source of a sign error that no test catches, because a flipped sign produces a perfectly valid result.

The value metrics are written as yields and not as multiples. That is not cosmetic. A multiple with a negative denominator flips sign. A company that loses money gets a negative price/earnings ratio, and in a sort from lowest first it ends up at the top, as the cheapest company in the universe. The yield form gives negative earnings a negative value that sorts at the bottom, where it belongs. The error is silent in the multiple form, and it affects precisely those companies a value filter is meant to weed out. The same applies in two other places. Leverage calculated on negative equity is not low debt, it is meaningless, and operating income divided by enterprise value is likewise meaningless when the company has more cash than market capitalisation. Both are set to missing rather than being read as a sign of quality.

The metrics are converted into rank percentiles rather than z-scores. Fundamental ratios have fat tails. A single return on equity of 400%, from a company with almost no book equity, moves both the mean and the standard deviation enough to change the scores of the entire block. A ranking cares about the ordering and not the distance, and is unaffected. A z-score is available as an alternative with symmetric clipping of 1% in each tail, so that the choice can be tested rather than assumed. Rank-based standardisation within sector follows the approach of Asness, Frazzini and Pedersen (2019).

Ranking takes place within sector group. Without it, the value block would in practice become a sector bet. Banks have structurally low price to book equity and software structurally high, and a pure cross-sectional ranking would fill the portfolio with financials in every period.

The choice of classification is justified by a measurement and not by a preference. None of the available classifications has history, so stability over time cannot be measured directly. What can be measured is

how much independent discretion each of them contains, measured against the industry code the company itself reports to the US authorities.

Classification	Categories	Code maps to several values
SIC sector	9	0.0%
Fama-French 48	48	1.1%
Morningstar sector	11	58.4%

The first two are in practice deterministic functions of a reported code and inherit its stability. Morningstar's sector classification is not: 58% of the codes map to several sectors, which affects 86% of companies. It therefore carries editorial discretion that can be revised without leaving a trace, and a revision would change a historical backtest.

The choice is the Fama-French twelve industries (Fama and French, 1997), derived from the industry code according to French's published definition (French, n.d.). It inherits the code's stability, it is the classification the literature uses, and it can be cited. The 48-industry classification would have been too fine: 48 groups across roughly 500 index members gives ten names per group, and the tail is worse. Among companies that have ever been index members, one group has a single name and two groups have two. A ranking within a group of one name measures nothing. With twelve industries the groups range between 7 and 100 names with a median of 36, and no group falls below five names at any rebalancing. The mechanism that merges groups that are too small has nevertheless been retained, because a universe may one day be narrower than today's.

The classification has a known bias that cannot be fixed with this data. The table it is drawn from is a current snapshot without history, with one row per company and no information about when the classification was set. A company that changed industry in 2007 carries its 2026 value in the 1999 rebalancing as well. The direction of the bias is worth noting. Sector neutralisation with hindsight classification places a company in the industry it ended up in, and compares it with the wrong peers in the early periods. The vendor offers no historical sector series, and the bias is therefore documented rather than corrected.

Companies with multiple share classes require a dedicated bridge. The index contains both classes for some companies, and two things go wrong if they are treated as two companies. The accounts are filed only under the primary class, so the secondary class would drop out as missing data even though the figures exist. And in an equal-weighted portfolio of 15 names, the company would be given 2/15 rather than 1/15, that is 6.7% of the portfolio allocated because of a capital structure and not a signal. The bridge finds the symbol that actually carries the fundamental data and deduplicates on that.

The bridge does not solve the case where the carrier of the accounting figures is itself not an index member. The company then drops out entirely. This is counted and logged per rebalancing, and the count produced a hit immediately. Twenty-First Century Fox files its accounts under a symbol that was not a member during the period, and the company therefore dropped out of seven rebalancings from December 2015 to December 2018. One name out of roughly 495 does not materially affect the result.

The point is that the assumption that the bridge always holds would have been wrong, and that this was discovered because it was counted.

Finally, all three blocks are required to have at least one computable metric. Without that requirement, a company lacking a value block would have been scored on two blocks out of three and compared with companies scored on three, that is, a comparison in which the weights mean different things from company to company. The alternative, assigning missing blocks the median score, is worse. The median sits in the middle of the pack, and a company would then gain from missing data every time the block would have dragged it down.

The block weights are equal. There is no measurement in this project that justifies an asymmetric weighting, and an optimised weighting would have been fitted to the same history that is later meant to evaluate it. Equal weight in the code turns out not to be equal weight in the portfolio, however, for a reason that only became visible once the score was measured. This is discussed in Chapter 4.7.

3.3 Statistical measures

The results are evaluated using a set of standardised measures established in quantitative finance research. Each measure addresses a specific dimension of strategy performance:

- Total return and CAGR (compound annual growth rate): measures absolute return over the period and the annualised growth rate. Central to assessing the strategy's ability to compound capital.
- Sharpe ratio (Sharpe, 1994): measures return per unit of total volatility. Allows comparison of strategies with different volatility levels.
- Sortino ratio (Sortino and Price, 1994): as Sharpe, but only downside volatility is penalised. Reflects the fact that upside swings are not undesirable.
- Maximum drawdown: the largest cumulative loss from peak to trough. Reflects the worst experienced decline and is critical for assessing whether the strategy is psychologically sustainable.
- Calmar ratio: CAGR divided by maximum drawdown. A measure of return relative to the worst decline.
- Profit factor: total gains divided by total losses. A measure of how much is made on winners relative to what is lost on losers.
- Win rate: the share of trades that are profitable. Less informative in isolation, but relevant in combination with the profit factor.

The results are compared against two benchmarks: passive buy-and-hold of a broad index ETF in the same market, and a classic 60/40 portfolio (60% equities, 40% bonds). These benchmarks are chosen because they represent realistic alternatives for the capital of HULC AS, and because they are what an active strategy must beat in order to justify its complexity.

For strategies operating on single stocks, these measures are supplemented by a factor regression. The strategy's excess return is regressed against established factor series for market, size, value, profitability,

investment and momentum (Fama and French, 2015; Carhart, 1997). The purpose is to distinguish between return that is due to exposure to known factors, and can be obtained more cheaply through index funds, and return that these factors cannot explain. A positive intercept that is not statistically significant is interpreted as the absence of a documented edge.

Turnover is reported two-sided in this report, that is, as the sum of purchases and sales measured against portfolio value. A two-sided turnover of 200% means the entire portfolio has been replaced once. The convention is stated here because it is not uniform in the literature, and because the two engines in the framework compute it differently: the portfolio engine two-sided and the ETF engine one-sided. Turnover figures from the ETF strategies in Chapters 4.3 to 4.5 must therefore be doubled before being compared with the figures for the single-stock strategies.

3.4 Validation and quality assurance

Several control mechanisms have been implemented to ensure that the results are credible:

- Unit testing: the unit tests cover critical components such as RSI calculation, cost modelling, no-look-ahead enforcement for both price and fundamental data, order simulation and report generation.
- Sanity checks along the way: early in the project a bug in the slippage modelling was identified when a backtest showed a 0% win rate over hundreds of trades, a statistically implausible result. This led to a revision of the model and illustrates how critical review of unexpected results is part of the method.
- Mutation testing of the control mechanisms: a test suite can be green on a rule it never exercises. To expose this, known errors were introduced into the code one at a time, and it was verified that the correct test did in fact turn red in each case. The errors were then removed and a clean state verified. The exercise was run in three rounds. First against the synthetic fixture suite, with six introduced look-ahead errors. Then against the suite that runs on the commercial dataset and on the portfolio layer, with thirteen introduced errors covering look-ahead, column confusion, settlement ordering, order filling, the commission floor, tax treatment, valuation of holdings and the wind-down of delisted positions. Finally against the multifactor scorer, with eleven introduced errors in sign, aggregation, filtering and grouping. None of the last eleven raise an error; all produce a valid but wrong result.
- Conservative assumptions: where there is room for interpretation, the conservative approach has been chosen. For example, when both the stop loss and the take profit fall within the same daily bar's range, it is assumed that the stop loss was hit first. This is pessimistic, but honest on daily data where intraday ordering cannot be determined.
- Explicit documentation of limitations: every result is documented with specific caveats relating to the period tested, the number of trades, and known methodological weaknesses.

The mutation testing revealed a pattern that matters more than the individual errors it found. On several occasions a test was green on a rule it did not exercise, and the common denominator was the same every time. The assertion was correct, and it was made against the correct object. What failed was that

the test scenario did not place the system in a state where the error could arise. Four cases illustrate the forms the failure takes.

Wrong date. The originally planned restatement test passed even when the version filter was removed entirely. It queried on a date at which the correction had not yet been published, and a version-blind implementation then returns the original value regardless. The rule would have been untested, with a green suite saying the opposite. The test was rewritten to query on a date after the correction became available.

Wrong resolution. The mutation that computed momentum on unadjusted rather than dividend-adjusted prices produced zero red tests. The test existed, and it compared momentum against the dividend-adjusted series. But the tolerance was 0.03, while the actual difference between the two series for the test company was 0.026. The test measured the right quantity at a resolution that did not distinguish right from wrong. It was rewritten to use four high-dividend-yield companies, and with the decisive assertion inverted: the momentum figure must lie far from the dividend-free series, not merely close to the adjusted one.

Wrong layer. The test of sector neutralisation showed that the scoring function neutralises correctly when it is given sector groups. It said nothing about whether the strategy actually passes groups. The mutation left the function alone and made the strategy stop calling it correctly, and the suite was still green. A test of a pure function tests the function, not that it is used.

Documentation versus code. The fourth is of a different kind, and it is the most serious. The year-end tax settlement routine took the year's realised gain out of the bucket and then sold down enough holdings to cover the tax liability. The gains from those sales were written back to the same year, a bucket that had already been emptied and is never read again. In the test scenario, over USD 200,000 of realised gains went untaxed. The module's own documentation described the correct behaviour precisely. The code did something else. No end figure looked wrong. The portfolio grew, the tax was paid, and the cash balance reconciled.

The consequence for how the work is done is the same in all four cases. None of them would have been found by code review. The first three are variants of a test that is too lax, where the code was correct and the test green. The fourth is worse, because both the documentation and the intention are correct, and only the code is wrong, in a way that makes the result better. A review of documentation against intention does not uncover it. All four were found by introducing the error deliberately and seeing whether the test noticed.

The round against the multifactor scorer showed the same thing at scale. Four of eleven mutations passed without a single red test on the first attempt, and four new tests had to be written against precisely those four errors. They have little in common in code and everything in common in form. None of them raises an error, none produces implausible figures, and all four would have produced a complete backtest result that looked entirely normal. One of them is worth singling out, because it is the most common way a factor model loses its integrity, namely filling missing data with the median. A company assessed on two blocks is then compared with companies assessed on three, and because the median

sits in the middle of the pack, the company gains from missing data every time the block would have dragged it down.

A separate finding concerns the design of the tests rather than their content. The mutation that computed market capitalisation on unadjusted prices turned only one of the two market-cap tests red. The broad test checks 400 companies against the dataset's own market-cap column with a threshold of 85% matches, and it passed, because large companies rarely reverse split. Only the narrow test turned red, the one that checks a single company chosen precisely because it is extreme. That company reverse split by a factor of 240, and with the wrong column its market capitalisation becomes 1.7 million instead of 415 million. The tail is not random. It is precisely the small and distressed companies that reverse split, and the error would consistently have made them look cheap. A value factor would have picked them systematically. An aggregate check with a threshold therefore cannot replace a concrete case chosen because it is extreme.

The exercise also showed that the number of red tests is a poor measure of how serious an error is. In the fixture suite, removing the version filter produced one red test, while filtering on period end rather than availability date produced nine. The two errors are equally destructive for a backtest result. In the suite against the commercial dataset, the same two errors each produce six red tests, that is, an entirely different ratio for the same pair of errors. The count says something about how many tests happen to touch a code path, not about what is at stake. Anyone who prioritises by coverage figures will misprioritise.

The mutations were chosen by the person who wrote the code, and therefore cover the errors he thought of. Errors in the field mapping itself, where a field turns out to measure something other than assumed, are caught neither by any of them nor by the suites more generally. The contract guarantees that the figure was known on the decision date, not that it is correct.

One final error was found by no test at all, but by reading two reports against each other. The portfolio engine matched the rebalancing date by exact equality against the trading calendar. If the date fell on a weekend or a public holiday, it did not exist in the calendar, and the rebalancing was skipped without a log line. This affected 16 of 55 dates over the period 1999 to 2026, that is 29%, and the affected periods therefore had a twelve-month interval rather than six. No end figure looked wrong. The error was discovered because the number of rows in a results file did not match the number of rebalancing dates stated in the text.

The fix rolls a rebalancing forward to the first following trading day. Forward and not backward, because the ranking is calculated as of the snapshot date. Executing it on the last trading day before that date would be trading on information that did not yet exist. The consequence is that eight of the sixteen dates, all year-ends, are now executed in January, so that the gain is realised in the following tax year. That is correct, since the trade actually takes place then. Both single-stock tests have been re-run after the fix, and the results in Chapters 4.6 and 4.7 come from the corrected engine.

Two things about this error are worth more than the error itself. It had a direction. The omitted rebalancings made the momentum strategy 1.16 percentage points better than it is, while barely

affecting the multifactor strategy. A random omission should not have had a systematic direction, and the cause has not been measured. And it reversed a conclusion. The effect of the ranking band in the multifactor strategy was measured at minus 0.12 percentage points with 39 rebalancings and plus 0.45 with all 55. A finding about an operating mechanism is thus also a finding about how often the mechanism is allowed to operate. The fix is now covered by a dedicated test, which deliberately uses a rebalancing date falling on a Saturday.

3.5 Limitations of the methodology

It is necessary to be explicit about what backtest results can and cannot tell us. The main limitations are:

First, the test periods vary substantially between strategies. The intraday and ETF strategies cover between six months and seven years, while the single-stock strategies are run over 27 years, from June 1999 to June 2026. In quantitative finance, 15 to 20 years is typically regarded as desirable for solid validation. That requirement is met for the single-stock strategies and not for the others. The caveat therefore applies not to the report as a whole, but to those tests where the period is in fact short.

Second, backtest results are by definition a simulation based on historical data. They do not reflect the psychological and operational factors that affect actual live trading: latency, partial fills, market impact, and not least the trader's own discipline under pressure. For this reason, the literature on backtest overfitting recommends discounting an observed Sharpe ratio before interpreting it as an expectation of future performance (Harvey and Liu, 2015; Bailey et al., 2014). The report does not state a quantified haircut, since its size depends on how many configurations have been tried.

Third, strategies that have no edge in the period tested should not automatically be discarded for all time. Market regimes change, and strategies may have an edge in the future even if they do not have one now, and vice versa.

Fourth, as a small Norwegian investment company we face natural resource constraints. Access to historical fundamental data has been solved for the US market through a commercial dataset, but corresponding coverage for European equities falls outside the project's budget. Nor do we have expensive backtesting platforms or full-time dedicated development capacity. What we have built represents what is practically feasible within these constraints, and that is reflected in the breadth over depth of our strategy selection.

Finally, if a specific strategy implementation has no edge, that is not the same as the entire approach being wrong. There may be variants, parameters, markets or time periods in which the strategy produces different results. What we can say is that we have tested specific specifications of the strategies under realistic conditions, and the results apply to those specifications.

There is also a limit to what the control mechanisms described in Chapter 3.4 can speak to. The fixture suite runs against synthetic data and proves that the mechanism works, not that a real dataset has the properties the mechanism presupposes. There is now in addition a dedicated suite that runs against the commercial dataset, and it shows that the dataset does deliver originally reported versions and genuine

availability dates. What remains unanswered is not the timing, but the values. The contract guarantees that a figure was known on the decision date, not that it is correct. A wrong sign, a wrong currency, wrong scaling or a key figure calculated on a different definition than assumed all pass unimpeded. Quality control of the values themselves is a separate task that has not been started, and constitutes an independent risk.

With these caveats stated, and with an honest acknowledgement of both what the method can and cannot tell us, we now turn to the experimental part of the report. The results that follow must be interpreted within these constraints, and the conclusion in Chapter 6 will summarise what we can actually say and what path forward this implies for HULC AS.

Two limitations apply to the multifactor strategy in particular. The sector classification is a current snapshot without history, as described in Chapter 3.2.3, and places companies in the industry they ended up in. The 180-day age limit for fundamental data is likewise not derived from the data. It is set because it covers two quarters with a normal publication lag, and it should be reconsidered against the actual reporting frequency in the dataset. Both are choices that are written down, so that they can be challenged.

4. Results

This chapter presents the results of the seven strategy tests carried out over the course of the project. The strategies are presented in the chronological order in which they were tested, and each section contains strategy-specific results, comparison against relevant benchmarks, and a brief interpretation of the findings.

4.1 Intraday RSI mean reversion (baseline)

The baseline strategy was tested on 33 US large-cap stocks over the period January 2024 to June 2024. The strategy opened positions in stocks with an RSI below 30 at the market open, with fixed 3% take-profit and 1.5% stop-loss bracket orders, and a time stop after 90 minutes.

The results were unambiguously negative, with a total return of -13.13%, a win rate of 33%, and 85% of positions closed by the time stop. That last figure is particularly revealing. It indicates that the expected mean reversion move largely did not materialise within the strategy's 90-minute horizon.

4.2 Filtered intraday variants (V2)

The combined v2 strategy (all five filters active) ended at -1.30%, a substantial improvement on the baseline but still negative. The isolated filter test showed that the gap size filter produced the most promising individual improvement (-0.41%), while ATR-based SL/TP and risk-based position sizing were actively harmful. Hypothesis 1 is falsified on solid empirical grounds for the conditions tested.

4.3 Factor ETF rotation for European markets

Three variants were tested over the period 2018-2024:

- V1 (pure relative momentum, default 6m/K=2): +41.4%, CAGR 5.07%, Sharpe 0.45, max DD -20.7%.
- V2 (Faber-style, SMA200 on each factor ETF): +3.1%, Sharpe 0.10, max DD -30.7%. The SMA filter was actively harmful.
- V3 (vol-weighted with a momentum tilt): +16.4%, Sharpe 0.24, max DD -29.6%.

By comparison, buy-and-hold IMEU returned +53.7% (Sharpe 0.45, DD -35.2%) and 60/40 returned +29.8% (Sharpe 0.42, DD -24.1%). None of the variants beat buy-and-hold on return, and only V1 matched it on Sharpe.

4.4 V1 parameter robustness

To examine whether the V1 result is robust to the choice of parameters, V1 was run across all 12 combinations of momentum lookback (3, 6, 9, 12 months) and top-K (1, 2, 3). The pre-specified definition of robustness required at least 75% positive Sharpe values, low dispersion, and no dramatic drawdown.

Results

V1 meets all three criteria: 12 of 12 have a positive Sharpe, the standard deviation (0.148) is lower than the mean (0.411), and the worst max DD (-33.7%) is within the threshold (-40.2%). Conclusion: V1 is ROBUST by this definition.

Main table of parameter sensitivity (Sharpe ratio):

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

The visualisation below (Figure 1) confirms the pattern in the table and illustrates the specific fragility at N=6, K=1. The heatmap shows that most parameter combinations produce Sharpe values between 0.34 and 0.63, with a clear outlier where N=6 is combined with a concentrated position (K=1).

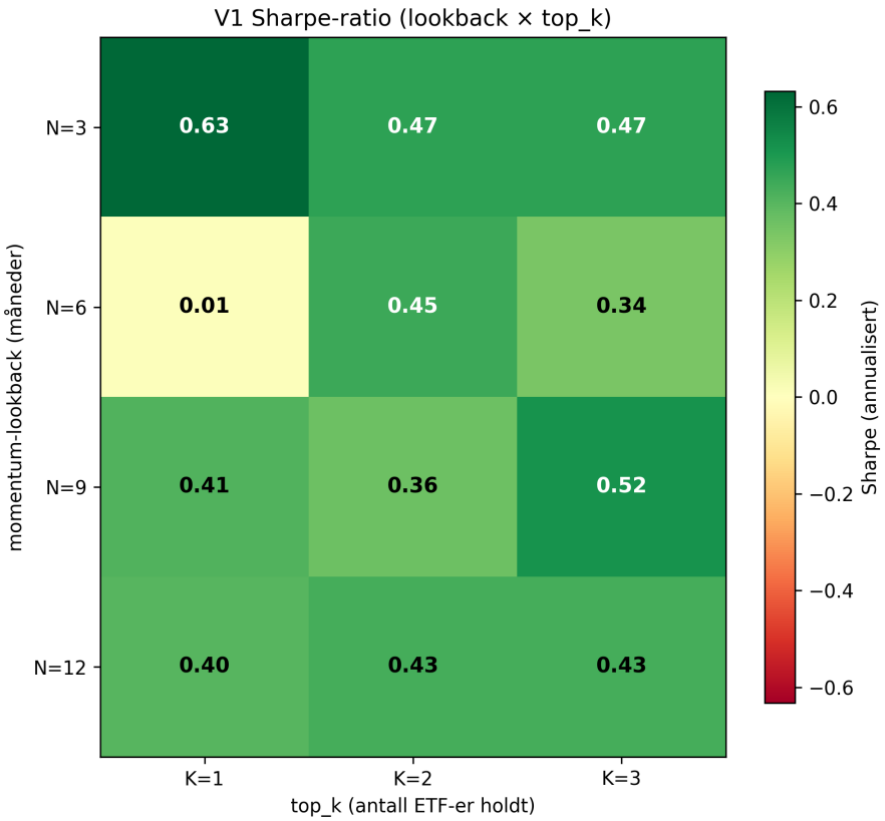

Figure 1: V1 Sharpe ratio across parameter combinations. The heatmap shows clearly that most combinations produce consistent results in the 0.34-0.63 range, but that N=6, K=1 stands out as an isolated weak cell (Sharpe 0.01). The result was verified as genuine fragility under a concentrated position, not an implementation bug.

Mean Sharpe across combinations: 0.41. By comparison, the IMEU benchmark has a Sharpe of 0.45 and 60/40 a Sharpe of 0.42.

In-sample / out-of-sample analysis

As an additional robustness analysis, the parameter grid was run separately on two halves of the period: 2018-2021 (IS) and 2022-2024 (OOS). The correlation between IS Sharpe and OOS Sharpe across the 12 parameter combinations is only 0.25, indicating that there was little to no reliable relationship between which parameters worked best in the first half and which worked best in the second.

Concretely, our original default choice (N=6, K=2) had an IS Sharpe of 0.82 but an OOS Sharpe of -0.09. Had we, at the end of 2021, selected the parameters that had worked best, we would have chosen a combination that did not work over the following three years.

Qualifying the robustness finding

The formal result (ROBUST by the definition) supports a weaker conclusion than it appears to at first glance. Three observations substantially qualify what 'robust' actually means:

The edge lies in drawdown, not in return. 12 of 12 V1 variants have a lower max drawdown than buy-and-hold, but only 4 of 12 beat the same benchmark on Sharpe ratio. The average CAGR (4.4%) is lower than the benchmark (6.33%). V1 reduces risk by giving up return.

The out-of-sample instability is the most revealing finding. An IS-OOS Sharpe correlation of 0.25 indicates that the strategy does not have a stable parameter configuration over time. We have no reliable basis for believing that the parameters that worked in the past would work in the future.

Our default configuration (N=6, K=2) was moderately lucky. Over the full period it is fine, but the N=6 row has the lowest average Sharpe (0.26 across top-K). N=3 had the highest average (0.52). The conventional 6-month momentum horizon was thus the worst on average in our sample.

The honest finding is that V1 is technically robust by the definition, but robustness here means 'does not collapse across parameters', not 'has a demonstrated edge over the benchmark'.

4.5 Absolute momentum baseline

After the V1 parameter robustness analysis had indicated that the V1 edge was marginal and unstable, the simplest possible trend-following strategy was tested as a baseline. If a simple switch between a broad index and government bonds could match or beat V1, that would confirm that the complexity of V1 (five factor ETFs with monthly rotation) added nothing beyond the underlying trend-following mechanism.

Strategy logic

The strategy follows Faber (2007) in its simplest form. On the last trading day of each month:

- If IMEU close > SMA(N): hold 100% IMEU
- If IMEU close <= SMA(N): hold 100% IEGA (European government bonds)

Two variants were tested:

- Variant A: SMA period of 200 days (classic Faber)
- Variant B: SMA period of 252 days (approximately 12 months)

The implementation follows the same architecture as the factor rotation strategies and uses the same engine, the same cost model and the same cached data. V1 was re-run in the same batch to ensure an identical basis for comparison, and ended at exactly +41.36%, matching the reference (consistency check passed).

Results

The main results over the period 2018-2024 are presented in the table below:

Strategy	Total return	CAGR	Sharpe	Sortino	Max DD	Calmar	Cost
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 (static)	+29.8%	3.80%	**0.42**	0.50	-24.1%	0.16	-
V1 pure momentum	+41.4%	5.07%	0.45	0.57	-20.7%	0.24	EUR 4,917

Analysis of regime shifts

A key finding is how the strategies handled the COVID crash in March 2020. Both variants moved to cash on 28 February 2020, which did cushion the initial decline. However, the strategy remained in cash throughout the V-shaped recovery that followed. Variant A re-entered the market on 30 November 2020, and Variant B later still. By that point most of the rally from the March low was already over - the strategy thus missed roughly 40% of return from the trough.

Figure 2 below illustrates the allocation history of both variants visually. A green background indicates equity exposure (IMEU), a grey background indicates a cash position (IEGA). The continuous grey block in 2020 clearly shows the COVID cash period, and the longer grey block in 2022 shows the bear market period. The difference between SMA 200 and SMA 252 is also visible. A shorter SMA produces more whipsaw effects (rapid regime shifts without clear direction), particularly in 2018.

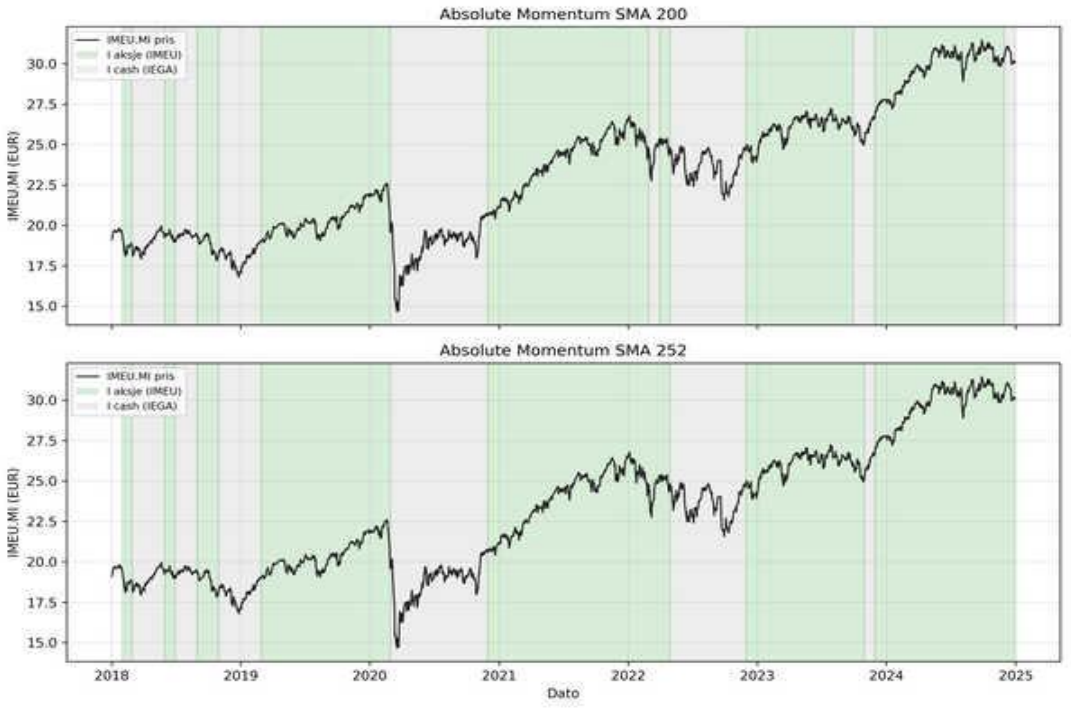

Figure 2: Allocation history for Absolute Momentum SMA 200 (top) and SMA 252 (bottom). Green background = position in IMEU, grey background = position in IEGA (cash). The black line shows the IMEU price. Long cash periods through 2020 (COVID) and 2022 (bear market) illustrate the strategy's tendency to miss V-shaped recoveries.

This is a classic weakness of trend-following strategies that rebalance monthly. The signals are too slow to capture V-shaped crashes. The strategy captures part of the downside but misses the recovery. In prolonged downturns (such as 2008 or the 2000-2002 dot-com crash) the strategy would probably have performed better, but those periods are not part of our sample.

In 2022 the variants behaved differently. SMA 252 (longer window, slower signals) stayed consistently in cash throughout the bear market and thereby avoided much of the downside. SMA 200, by contrast,

was exposed to whipsaw effects, switching between cash and equities several times without clear direction.

Cost analysis

The absolute momentum strategies have substantially lower transaction costs than V1, as expected. Variant A incurred total costs of EUR 3,019 and Variant B EUR 2,396 over the period, against V1's EUR 4,917. Variant B is thus roughly half as costly as V1, but delivers substantially lower returns.

Comparison against V1

The result is uncomfortable but honest: the complexity of V1 does in fact beat the simple switch. V1 returned +41.4% while the AbsMom variants returned 26-29%. The Sharpe ratios are also higher for V1 (0.45 vs 0.35-0.37). V1 loses to SMA 252 only on max drawdown (-20.7% vs -18.0%). This comparison must at the same time be read together with the finding in Chapter 4.4: V1 beats the simple switch over the full period, but without a parameter configuration that is stable out of sample. The two findings are not in conflict, but they constrain each other, and are reconciled in Chapter 4.8.

This qualifies an important observation. Trend following in itself is not the problem. It is trend following at index level (absolute momentum) that is weaker than trend following at factor level (V1 pure momentum). V1 always holds a position in the most attractive factor ETFs, whereas absolute momentum goes entirely to cash and thereby misses the recovery in a V-shaped crash.

Comparison against a static 60/40

The single most important finding from this test is the comparison against the 60/40 portfolio. A static 60/40 allocation (60% IMEU, 40% IEGA, rebalanced monthly) delivered +29.8% with a Sharpe of 0.42 and a max DD of -24.1%.

60/40 thus beats both AbsMom variants on:

- Total return (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)

And it does so without timing, without monthly regime decisions, and without transaction costs associated with regime shifts. A boring static mix thus beats both timed variants on almost every relevant measure. The only measure on which 60/40 loses is max drawdown, where SMA 252 has the lowest of all (-18.0% vs -24.1%).

The comparison is unambiguous. The entire SMA apparatus, with all its theoretical appeal and complexity, failed to earn its own existence over this period. If HULC AS wants lower drawdown than buy-and-hold, a static 60/40 dominates both SMA variants on almost every measure.

Caveats

It is important to emphasise that 7 years containing a single V-shaped crash is structurally the worst possible regime for monthly trend following. The historical reputation of the Faber strategy was earned

in 2008 and 2000-2002, which were prolonged downturns in which monthly rebalancing was fast enough to capture most of the downside without missing the upside. Such periods are not part of our sample.

The conclusion regarding the absolute momentum strategy is therefore suggestive rather than conclusive. The strategy performs poorly in V-shaped crashes, but may potentially perform better in prolonged downturns. The 60/40 comparison, however, is independent of the shape of the crash, and that comparison argues clearly against adopting the absolute momentum strategy.

Conclusion on absolute momentum

The recommendation is not to adopt either AbsMom variant for HULC AS. If the company wants lower drawdown than buy-and-hold, a static 60/40 portfolio dominates them on almost every measure. Taken together across factor rotation and absolute momentum, we find that neither intraday mean reversion, factor rotation nor index trend following has a reliable edge in this sample, and that a boring static allocation matches or beats them all.

4.6 Cross-sectional momentum in US single stocks

Cross-sectional momentum over 12 months excluding the most recent month, top 15 equal-weighted, semi-annual rebalancing. S&P 500 members with historical membership, 55 rebalancings from 30 June 1999 to 30 June 2026. Costs, T+1 settlement and 22% tax as described in Chapter 3.1. The benchmark is SPY buy-and-hold with the same cost treatment for the initial position. The strategy is included as a control group. It shows what momentum alone delivers on the universe, so that any contribution of the multifactor strategy beyond price can be isolated in Chapter 4.8.

	Test A 50k	Test A 100k	S&P 500
Total return	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
Max drawdown	−86.23%	−85.64%	−55.19%
Calmar	0.050	0.055	0.153
Hit rate	48.00%	48.52%	—
Turnover, annual (two-sided)	284.26%	284.26%	0%
Tax paid (USD)	3,075	7,190	0
Terminal value (USD)	19,929	44,240	57,174 / 114,492

Starting capital is USD 6,372 at NOK 50,000 and USD 12,743 at NOK 100,000, converted at 7.8471. The benchmark is not beaten on any metric. The strategy is 4.1 percentage points below the index in annual return, with less than half the Sharpe ratio and a drawdown 31 percentage points deeper.

Costs are decomposed by component. Commission is almost identical in dollar terms at the two capital levels, USD 1,313 against USD 1,367, because the number of orders is almost the same and the one-dollar floor binds in both. It therefore accounts for twice as large a share of turnover at the lower level, 0.296% against 0.142%. Total cost drag is 1.53 percentage points per year at NOK 50,000 and 1.10 at NOK 100,000. That is the commission floor measured directly, and it is the basis for the assessment in Chapter 3.1.3.

Cost	50k, USD	50k, pp/yr	100k, USD	100k, pp/yr
Commission	1,313	0.840	1,367	0.404
Spread	89	0.057	193	0.057
Slippage	994	0.636	2,150	0.635
Total	2,396	1.533	3,710	1.095

The ranking band, that is, the rule that a held position is retained as long as it remains within the top 23 even though the portfolio holds only 15 names, adds 0.15 percentage points of return at NOK 50,000 and 0.24 at NOK 100,000. Turnover falls from 313% to 284%, and 63 positions were retained solely because they were inside the buffer. Notably, the band here increases the tax paid, from USD 2,625 to USD 3,075, while also increasing the return. That is the opposite of the mechanism one would expect. A plausible explanation is that the variant without a band realises more losses, which do not trigger tax, but the mechanism has not been measured and should not be presented as though it had been.

The distribution across five-year sub-periods is the same story told six times.

Period	Test A	S&P 500	Excess return
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%

* Incomplete sub-period of 2.09 years. It is included rather than omitted, since an omitted sub-period is a silent sampling decision.

The strategy beats the benchmark in 2 of 6 sub-periods, and both are the two most recent. That is worth noting against the overall picture. The strategy that loses clearly over the full history is the one that has been best in recent years.

The NAV path explains why the return is low despite the strategy performing very well in some periods. The portfolio rose from USD 6,335 to USD 13,240 by 23 March 2000, that is, a doubling in nine months,

and then fell. The trough came on 20 November 2008 at 86.2% below the peak, that is USD 1,824. The old peak was not regained until 27 June 2025, more than 25 years later.

Stress episode	Window	Test A	S&P 500
Dot-com	2000-03 → 2002-12	−74.5%	−41.6%
Financial crisis	2007-10 → 2009-03	−75.0%	−54.6%
Covid	2020-02 → 2020-04	−35.5%	−30.8%

The holdings confirm that the strategy did exactly what it is built to do. In June 2000 it owned AMD, Sun, Micron, Oracle and Siebel, that is, the dot-com names right before the crash. In June 2008 it owned energy and coal right before the commodity collapse. In December 2020 it owned growth names right before the 2021 correction. Momentum buys what has risen, and concentration in 15 names turns each of the three episodes into a full collapse rather than a setback. The pattern is known in the literature as a momentum crash (Daniel and Moskowitz, 2016). The conclusion is not that the strategy is misimplemented, but that the strategy is what it is.

The tax is low in absolute terms, USD 3,075 at NOK 50,000. That is no advantage. The strategy realised large losses in 2001, 2002 and 2008, and carried them forward in 23 of 28 years. Low tax here is the consequence of losing money.

4.7 Multifactor (QVM) in US single stocks

The multifactor strategy ranks on a composite score of quality, value and momentum with equal weight on the three blocks, within sector group, as the construction is described in Chapter 3.2.3. Top 15 equal-weighted, semi-annual rebalancing, 55 rebalancings from 30 June 1999 to 30 June 2026. Everything else is identical to Test A: the same universe, the same costs, the same settlement, the same tax and the same ranking band. The configuration is imported from Test A rather than restated, so that the two cannot drift apart.

The criteria were defined in advance, before the test was run. The strategy must beat both the passive benchmark and cross-sectional momentum after costs and tax, and have statistically significant alpha in the factor regression. The result is assessed against them at the end of this section.

	Test B 50k	Test B 100k	S&P 500
Total return	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
Max drawdown	−58.00%	−57.96%	−55.19%
Calmar	0.110	0.113	0.153
Hit rate	56.97%	56.97%	—

Turnover, annual (two-sided)	251.66%	251.80%	0%
Tax paid (USD)	6,934	14,667	0
Terminal value (USD)	34,027	70,850	57,205 / 114,554

The benchmark is not beaten. The multifactor strategy is 2.1 percentage points below the S&P 500 in annual return, with a lower Sharpe, a lower Calmar and a marginally deeper drawdown. That is the headline result. The rest of this section explains how it came about, not why it was really better than it looks.

The cost drag is 0.75 percentage points per year at NOK 50,000 and 0.61 at NOK 100,000, that is, less than half of Test A's despite almost identical turnover. The reason is that the strategy has more capital to trade with throughout the period. It does not erode down to below USD 2,000 along the way, and the commission floor therefore bites far less. The cost level is thus as much a function of how the strategy performed as of how it trades.

Cost	50k, USD	50k, pp/yr	100k, USD	100k, pp/yr
Commission	1,214	0.272	1,279	0.137
Spread	222	0.050	466	0.050
Slippage	1,886	0.423	3,962	0.424
Total	3,321	0.745	5,707	0.611

The ranking band adds 0.45 percentage points of return at NOK 50,000 and 0.43 at NOK 100,000, that is, the same order of magnitude and the same sign as in Test A. 72 positions were retained solely because they were inside the buffer. This figure is a correction. With the earlier engine, which skipped 16 of 55 rebalancings, the band was measured at minus 0.12 percentage points, and the report concluded that it did the opposite of what it did in Test A. With all rebalancings in place it behaves as expected. The point generalises beyond this table: a finding about an operating mechanism is also a finding about how often the mechanism is allowed to operate.

The distribution across sub-periods is the most troubling part of the whole result.

Period	Test B	S&P 500	Excess return
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%

* Incomplete sub-period of 2.09 years.

The strategy beats the benchmark in 2 of 6 sub-periods, and those two are the first two. The pattern is clear. The entire excess return was earned before 2009, and every single sub-period after the financial crisis loses to the index. A strategy that worked in the first two thirds of the sample and not in the last is not a strategy with an edge. It is a strategy with a period.

The NAV path runs from USD 6,339 to USD 34,027, with a peak of USD 34,503 on 28 July 2026. The worst drawdown is 58.00%, from USD 18,293 on 13 July 2007 to USD 7,684 on 3 March 2009, and the old peak was not regained until 19 March 2014, that is, six and a half years under water.

Stress episode	Window	Test B	S&P 500
Dot-com	2000-03 → 2002-12	−5.5%	−41.6%
Financial crisis	2007-10 → 2009-03	−57.1%	−54.6%
Covid	2020-02 → 2020-04	−45.6%	−30.8%

The three episodes are explained by the holdings, and they point in different directions. In June 2000 the portfolio owned railways, chemicals, insurance, industrials and a newspaper, not the dot-com names. It is the value and quality blocks overriding momentum. A company without earnings gets a negative earnings-to-market-cap figure and falls to the bottom of the value block no matter how strongly its share price has risen. Dot-com is the one episode the strategy handled well, and it explains almost the entire excess return over 27 years.

In June 2008, by contrast, it owned energy and materials together with four insurance companies, that is, companies that were cheap on accounting figures right before both the commodity collapse and the financial crisis hit precisely those sectors. The value block pointed into the crisis rather than away from it, and the drawdown was deeper than the index's. In December 2020 it owned homebuilders, insurance and healthcare, fell 46% against the index's 31% in the COVID crash, and did not recover it in the following years.

The tax is USD 6,934 at NOK 50,000, that is, twice as high as in Test A in absolute terms. The reason is simple: the strategy made more money. Test A carried losses forward in 23 of 28 years, Test B in 10. That is not tax optimisation, it is a strategy that loses less often.

Qualification is the one step Test B has that Test A does not. A company must have sufficiently recent values in all twelve fundamental fields. The distribution across the 55 rebalancings shows that the requirement is mild in practice.

Step	Median	Min	Max
Index members	500	500	505
Qualified	496	466	500
Excluded: missing fields	5	0	20
Excluded: data older than 180 days	1	0	16
Excluded by the strategy	1	0	4

Ranked	495	466	500

The sector classification produced twelve groups at every single rebalancing, so the mechanism for merging small groups was never triggered. The 64 strategy exclusions in total are companies where an entire block was missing, in practice negative equity or negative enterprise value. The portfolio used 310 unique names across the 55 rebalancings. The survivorship bias arising from the exclusions is discussed in Chapter 3.2.2.

The final finding concerns the construction itself, and it only became visible once the score was measured. The median score among the selected names is systematically highest in the momentum block.

Block	Mean	Standard deviation	Min	Max
Quality	0.213	0.056	0.051	0.329
Value	0.239	0.057	0.103	0.335
Momentum	0.389	0.037	0.278	0.450

The reason is not that momentum is weighted more heavily. The weights are equal. The reason is that momentum is a single metric while quality and value are averages of three and four metrics respectively. An average of several rankings pulls towards the middle, so a block with more metrics has less dispersion than a block with one. Measured across all ranked names, each individual metric is equally dispersed, with a standard deviation of 0.289, while the blocks are not: roughly 0.19 for quality, 0.20 for value and 0.289 for momentum. On dispersion alone, that corresponds roughly to an effective weight of about 42% on momentum against 29% each on quality and value.

Equal weight in the code is therefore not equal weight in the portfolio, and no one decided that it should not be. This is a construction flaw, and it is worth testing in the robustness sweep rather than fixing after the fact. Normalising the block scores before weighting would make the weights real. The flag that does this exists in the code, but it was added after the test had been run, and it is therefore off by default. Changing the construction after seeing the result would have turned the robustness sweep into a post-hoc rationalisation.

Measured against the criteria, the strategy beats cross-sectional momentum on every single metric, it does not beat the passive benchmark on any metric, and the factor regression has not been run. The one criterion that is settled against the strategy is the one that weighs most heavily. A strategy that loses to an index fund after costs and tax, over 27 years, cannot be defended on the grounds that it beat another active strategy. The factor regression remains outstanding and would be able to show whether what is left of the return is exposure to known factors, but it cannot change the headline result.

Robustness and out-of-sample

A single result says little about whether a strategy is worth anything. The question is whether the result is a property of the construction or of the particular parameter choice that happened to be run. This has

been tested with a sweep across 30 combinations. Five weighting sets, namely equal weight, quality only, value only, momentum only, and quality plus value, combined with three portfolio sizes of 10, 15 and 25 names, with and without sector neutralisation. The same period, the same costs, the same tax and the same settlement as in the tests above. The metrics are computed once per rebalancing date and rescored for each combination, so that the fundamental lookup is identical across the grid.

The criteria were written into the code before the first run and have not been changed afterwards. Robust here means that the strategy does not collapse within the parameter space, not that it has an edge. The wording is inherited from the sweep in Chapter 4.4 and is just as necessary here.

Criterion	Measurement	Outcome
At least 75% have positive Sharpe	30 of 30	Met
Standard deviation lower than the mean	0.065 vs 0.422	Met
No drawdown more than 5 pp worse than the benchmark	Worst −84.89%	Not met
At least 75% with lower volatility than Test A	29 of 30	Met

Three of the four criteria are met. The third is not, and not in some corner of the grid. Half the grid, 15 of 30 combinations, has a drawdown worse than the 60.19% threshold. The worst are momentum alone without sector neutralisation, in practice Test A, but all six value variants, three of six quality variants and two equal-weight combinations also breach the limit. The depth of the declines is not a property of one unfortunate parameter choice. The overall verdict is that the grid is not robust.

The volatility figures in this table are annualised from daily returns. The same applies to the decomposition tables at the end of Chapter 4.8. The volatility drag table early in Chapter 4.8, by contrast, annualises from monthly returns and therefore reports lower figures for the same strategies. Both are valid, and the daily figure is higher because it captures intraday swings that are smoothed out over a month. The comparison in the fourth criterion uses the same estimator on both sides.

An extract from the grid, sorted by Sharpe, shows the pattern. Over the same period the benchmark has a CAGR of 8.46%, a Sharpe of 0.519 and a maximum drawdown of 55.19%.

#	Weights	K	Sector	CAGR	Sharpe	Max DD	IS	OOS
1	Q+V	25	flat	9.27%	0.514	−49.59%	0.569	0.451
2	Q only	25	flat	8.84%	0.505	−52.36%	0.351	0.674
3	Q+V	25	sector	8.50%	0.491	−56.18%	0.480	0.501
4	Q only	15	flat	8.88%	0.490	−56.13%	0.371	0.637
5	Q+V	15	sector	8.77%	0.490	−54.88%	0.474	0.510
6	V only	15	flat	9.84%	0.490	−71.99%	0.406	0.583
7	Q+V	10	sector	9.03%	0.486	−51.09%	0.523	0.443
20	Equal (Test B)	15	sector	6.40%	0.394	−58.00%	0.315	0.478

29	M only	15	flat	4.69%	0.304	−84.89%	0.007	0.674
30	M only	10	sector	4.47%	0.300	−61.56%	−0.044	0.657

Two patterns recur. More names is better: 25 names produces the highest average Sharpe in all five weighting sets and the lowest drawdown in four of five, and the ordering 25 ahead of 15 ahead of 10 holds strictly in four of five. Concentration in fifteen names costs, and it costs most in volatility. And the momentum block is the weakest. The seven top combinations contain no momentum block at all. Only in eighth place does a combination with momentum appear.

The combination Test B actually ran ranks 20th of 30. It is not a lucky point in the grid, it is a mediocre one. Nineteen combinations did better, and they share two features: less momentum and more names.

It is tempting to read the best cell as a discovery. Quality plus value with 25 names without sector neutralisation delivers 9.27% annual return against the benchmark's 8.46%, with a drawdown of 49.6% against 55.2%, and is thus the only configuration in the whole report that beats the index. That reading does not hold, and the sweep's own figures say why.

The split into in-sample and out-of-sample falls at the turn of the year 2012/2013, on the NAV path from a single continuous run rather than two separate backtests. The correlation between Sharpe in the first and second halves, across the grid, is minus 0.72. By comparison it was plus 0.25 for the factor rotation in Chapter 4.4, and that was called the single strongest piece of evidence against that strategy at the time. Which parameters worked in the first half said almost nothing about which worked in the second. Minus 0.72 is not nothing. It is worse. The parameters that performed best in the first half performed systematically worst in the second.

The mechanism is visible in the table and is not noise. Momentum alone has a Sharpe of around zero in the first half, from minus 0.04 to 0.13, and between 0.66 and 0.76 in the second. Value alone and quality plus value show the opposite pattern. Value and quality led from 1999 to 2012, momentum led from 2013 to 2026. The grid therefore largely measures which factor regime each half belonged to, not which parameters are good. The consequence is concrete. Had one chosen a weighting set on data through 2012, one would have chosen quality plus value, and obtained the weakest half of the outcomes thereafter. The best combination in-sample ranks 29th of 30 out-of-sample.

The sweep also answers the question left open in Chapter 4.8. Is the low volatility a property of the construction or of the parameter choice? The fourth criterion answers that it is a property of the construction. 29 of 30 combinations have lower volatility than pure momentum, and the one that does not is momentum alone with ten names without sector neutralisation, that is, the combination closest to Test A to begin with. That exception is worth noting, because it points to which mechanism is doing the work. The decomposition at the end of Chapter 4.8 shows that it is mainly the sector neutralisation and not the fundamental blocks that lowers volatility.

The construction flaw in the block weights has also been tested. Normalising the block scores before weighting, so that equal weights become equal weights, gives six better and six worse outcomes across

twelve comparisons, with no pattern as to which. The weakness is real and without practical consequence on this grid. That is worth writing down precisely because the opposite was expected.

No single combination should be promoted on this basis. Quality plus value with 25 names looks best, but it was selected after the fact in a grid where selection on the first half demonstrably points the wrong way. Picking it would be doing exactly what the sweep demonstrates one should not do.

4.8 Summary comparison

This section answers the question the report was written to address. Do the fundamental blocks add anything beyond price alone? Test A and Test B are re-run in the same process and on the same price history rather than comparing two sets of result files, and the benchmark is scaled once and used for both.

The first question must be whether the comparison measures two different portfolios at all. If the selections are in practice the same names, one is comparing two variants of the same portfolio and calling them two strategies. That is not the case here. The two portfolios share on average 2.15 of 15 names, that is 14.3%, with a median of 2 and a range from 0 to 5. Six of the 55 rebalancings have no name in common at all. The overlap is moreover stable throughout the history, between 1.95 and 2.50 on average per decade. The momentum block accounts for one third of the multifactor score, and a name must also be good on quality and value to be included. On average, two out of fifteen manage that.

Test B beats Test A on every single metric: 6.40% against 4.32% in annual return at NOK 50,000, 0.394 against 0.293 in Sharpe, a drawdown 28 percentage points shallower, and lower turnover. Neither beats the benchmark on any metric. The difference can be split in two by re-running both without costs and without tax.

Capital	A gross	B gross	A net	B net	Allocation	Friction
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

Eight ninths of the difference is allocation, that is, what the strategies owned. Friction also favours multifactor, but only marginally. Lower turnover saves costs, while higher returns generate more tax, and the two almost offset each other. Worth noting separately is that friction costs Test A 2.5 percentage points per year and Test B 2.1. Both strategies thus spend around a quarter of their gross return on costs and tax.

The next table is the most important in the report, because it shows that the difference is not where one would expect.

	Mean per month	Annualised mean	Annualised vol	CAGR	Volatility drag
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

Multifactor does not earn more than pure momentum. It loses less along the way. At the NOK 100,000 level the arithmetic means are 8.06% for Test A and 8.04% for Test B, that is, identical. The entire CAGR difference of 2.1 percentage points is the difference in volatility drag, 26% annual volatility against 18%. A 50% loss requires 100% to recover, and Test A had three such episodes. What creates the lower volatility is decomposed in the next section, and the answer is not what one would expect.

That makes the result both better and worse than it looks. Better, because lower volatility is a real and valuable property that does not vanish as readily as an excess return does. Worse, because there is no sign that the fundamental blocks find companies that rise more. They find companies that fall less, and the index does both better than either strategy.

The question is then whether the difference is larger than noise. A paired test on monthly returns over 326 observations gives the answer.

Pair	Correlation	Mean per month	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

The monthly excess return of Test B over Test A is indistinguishable from zero. This is not a marginal p-value. It is 0.96, that is, as close to zero effect as a test can come. This is not a contradiction of the 2.1 percentage point CAGR difference. The test measures the arithmetic mean, and there the two are equal. The difference in terminal value comes from the second moment, not the first, and a t-test on the mean does not see it.

That Test B against the benchmark is not significant either, with $p = 0.50$, says something about what 27 years actually affords. With a monthly standard deviation of 3% in the difference, far more than 326 observations are required to demonstrate an excess return of a couple of percentage points. The absence of significance is not proof that the strategies are alike. It is a reminder that this sample cannot settle the question in either direction.

The holdings in the three stress episodes show how little the two strategies have in common. In June 2000, Test A owned the semiconductor and networking names right before the crash, while Test B owned railways, chemicals, insurance and a newspaper. One name in common. In June 2008 and December 2020 there were no names in common at all. In 2008 both lost, because Test B had bought the other cheap names in the same crisis. In 2020, Test A owned growth names and Test B owned homebuilders and healthcare, and there Test A won clearly in the following years.

The conclusion has four parts. Multifactor does add something real. The portfolios are different, and Test B is better than Test A on every metric. But the contribution is risk reduction and not excess return, and the difference is not statistically distinguishable from zero on monthly data. The distribution over time is worse than the level. Test B wins the first two five-year periods and loses the last two, so the sign of what multifactor contributes depends on which decade one asks about, and the period that decided the matter, dot-com, is a single event. And the risk reduction comes mainly from the sector neutralisation, not from the fundamental blocks, as shown in the decomposition below.

Where the improvement actually comes from

The comparison above has a weakness that must be dealt with before it can be interpreted. Test A and Test B differ in two respects at once, namely the signal and the sector neutralisation, and the entire difference has so far been attributed to the former. It is not obvious that this is correct. A dedicated run separates the two terms by adding one mechanism at a time.

The volatility figures in this section are annualised from daily returns, cf. the note in Chapter 4.7, and are therefore higher than the figures in the volatility drag table above. The intermediate variant is Test A with ranking within sector group and nothing else changed. The same membership list, the same merging of share classes, no qualification on fundamental data. It is run with the same costs, tax and settlement as the other two.

Step	Volatility	Max drawdown	Sharpe	CAGR
Pure momentum, no sector constraint	31.63%	−86.23%	0.293	4.32%
With sector constraint	24.00%	−58.23%	0.329	5.11%
With quality and value added	21.91%	−58.00%	0.394	6.40%

The sector constraint alone takes volatility down by 7.63 percentage points and the drawdown up by 27.99 percentage points. The fundamental blocks, layered on top, take volatility down by a further 2.09 percentage points and the drawdown up by 0.24. The sector constraint thus accounts for 99% of the improvement in drawdown and 79% of the improvement in volatility.

That is a material correction to the picture the rest of the chapter paints. The risk reduction that distinguishes multifactor from pure momentum comes mainly from spreading the portfolio across industries, not from selecting companies on accounting figures. A concentrated fifteen-name momentum portfolio ends up in one industry at a time, and that is what produces the 86 percent declines.

The question then becomes whether it is the actual industry classification that works, or merely spreading the portfolio across twelve groups whichever they are. This has been tested by shuffling the group labels at random, with five seeds, and otherwise running identically.

Variant	Volatility	Max drawdown	Sharpe	CAGR
Without sector constraint	31.63%	−86.23%	0.293	4.32%

Permuted groups, mean of five	29.37%	−81.03%	0.319	5.16%
True industry classification	24.00%	−58.23%	0.329	5.11%

Random groups reproduce 19% of the effect on drawdown and 30% of the effect on volatility. Four fifths of the reduction in declines thus requires the groups actually to be industries. Diversification in itself explains part of it, but not most of it, and that is an argument that sector neutralisation does something real and does not merely split the portfolio into more pieces.

The third term is the qualification requirement. A company must have current values in all twelve fundamental fields to be ranked at all, and that removes companies from the universe before any signal is computed. That term can be separated from the score itself by running momentum alone on the qualified universe, that is, with the filter on and the signal off.

Step	What is added	CAGR	Max drawdown	Sharpe
A	—	4.32%	−86.23%	0.293
A2	Sector constraint	5.11%	−58.23%	0.329
A2 qualified	Fundamental data as filter	5.75%	−58.61%	0.353
B	Fundamental data as signal	6.40%	−58.00%	0.394

Of the 2.08 percentage point difference in annual return between the two tests, 0.79 comes from the sector constraint, 0.63 from the qualification filter and 0.65 from the multifactor score itself. Less than a third of the advantage is therefore due to the signal the hypothesis is about.

The filter term is moreover not a neutral restriction, and the direction of the bias has already been measured in Chapter 3.2.2. Companies lacking fundamental coverage are overrepresented among those that disappeared from the index, with a relative risk of 2.73. The filter prevents the strategy from owning them. The 0.63 percentage points from the filter term are therefore not pure results, but partly a measurement of a known bias with a known direction. The magnitude of the bias is small, as shown in the same chapter, but the term cannot be read as a property of the strategy.

The decomposition has a practical consequence that is worth more than the rest of the comparison. Sector neutralisation requires no fundamental data. It requires only an industry classification, which is available free of charge, and it delivers almost the entire reduction in declines. The commercial dataset, which is by far the largest cost in the whole setup, pays for the remaining 1% of the drawdown improvement and 0.65 percentage points of annual return, part of which is again a filter effect.

Finally, it is worth noting that Test A is the only one of the two that beats the benchmark in any period after 2009. The strategy that loses clearly overall is thus the one that has been best most recently. That is an argument for caution in reading a ranking between two strategies as a durable relationship.

The reconciliation against the ETF strategies in Chapters 4.4 and 4.5 remains. The two findings there, that V1 beats the simple index switch on most measures while its parameter configuration is unstable

out of sample, still stand, and what they jointly imply for the choice of strategy belongs here and in Chapter 5. Note that the turnover figures for the ETF strategies are reported one-sided and must be doubled before being compared with the figures above, cf. Chapter 3.3.

The question of whether low volatility is a property of the construction or of the parameter choice is answered in the robustness sweep at the end of Chapter 4.7. It is a property of the construction, since 29 of 30 combinations have lower volatility than pure momentum. Which part of the construction is answered in the decomposition above, and the answer is the sector constraint. The same sweep shows at the same time that the grid is not robust on drawdown, and that the correlation between the first and second halves is minus 0.72. The factor regression described in Chapter 3.3 remains outstanding, and would be able to show whether what is left of the return is exposure to known factors.

5. Discussion and synthesis

Five strategy families have been tested: intraday RSI mean reversion, filtered variants of it, factor ETF rotation in European markets, absolute momentum as trend following, and two single-stock strategies in the US market. None of them delivered risk-adjusted returns exceeding a passive benchmark portfolio after costs and tax. That is the answer to the extended research question in Chapter 1.2, and it is clear enough that it should be stated before the nuances.

5.1 What the five families showed

Hypothesis 1 is falsified. The baseline variant of intraday RSI returned minus 13.13% over the period, with 85% of positions closed on the time stop, that is, without the expected move materialising at all. Filtering lifted the result to minus 1.30%, but the sign did not change, and two of the five filters were actively harmful. The falsification applies to the conditions tested, not to the idea of mean reversion as such.

Factor rotation was the first strategy to pass a formal robustness test, and it is at the same time the best illustration of how little that means. All twelve parameter combinations produced a positive Sharpe, dispersion was low, and no combination had a dramatic drawdown. But the edge lay in drawdown and not in return: 12 of 12 variants had shallower declines than buy-and-hold, while only 4 of 12 beat the same benchmark on Sharpe, and average return was lower. And the correlation between the first and second halves of the period was 0.25.

Absolute momentum was tested as the simplest possible baseline, to see whether the complexity of the rotation added anything. It did. The rotation beat the simple index switch on return and on Sharpe. But both lost to buy-and-hold, and the weakness of trend following was clearly exposed in 2020, where the strategy went to cash in February and did not return until November, that is, after most of the recovery was over.

The two single-stock strategies are the most comprehensive tests in the report, and they are the only ones run over 27 years with point-in-time data. Pure momentum returned 4.32% per year against the

benchmark's 8.46%, with a drawdown of 86%. Multifactor returned 6.40% with a drawdown of 58%. Multifactor beats pure momentum on every metric. Neither beats the benchmark on any metric.

5.2 The pattern across the families

Three features recur in all five families, and they are more interesting than the individual results.

The first is that where a strategy had something to offer, it lay in risk reduction and not in excess return. Factor rotation reduced declines and gave up return. Multifactor did not earn more than pure momentum. The arithmetic means are identical, 8.06% against 8.04% at NOK 100,000, and the entire difference in terminal value arises because volatility is 18% against 26%. That is a real and valuable property, but it is a different property from the one the hypothesis was about.

And the risk reduction does not come from where one would expect. The decomposition in Chapter 4.8 separates the sector neutralisation from the fundamental blocks, and the sector constraint alone accounts for 99% of the improvement in drawdown and 79% of the improvement in volatility. Of the 2.08 percentage point advantage in annual return, 0.79 comes from the sector constraint, 0.63 from the qualification filter and 0.65 from the multifactor score itself. What works, then, is mainly that the portfolio is spread across industries. A concentrated fifteen-name momentum portfolio ends up in one industry at a time, and that is what produces the 86 percent declines.

The second is that parameter choices did not transfer through time. Factor rotation had a correlation of plus 0.25 between the halves, and that was at the time called the single strongest piece of evidence against the strategy. Multifactor has minus 0.72. That is not an absence of relationship, it is an inverse relationship. The parameters that worked best in the first half worked systematically worst in the second. In both cases a choice made on half the history would have pointed the wrong way. That is the single strongest argument in the entire report against selecting a configuration on historical data.

The third is that the cost and tax model is a larger factor than the choice of strategy at the capital levels at which HULC AS operates. Friction costs pure momentum 2.5 percentage points per year and multifactor 2.1, that is, around a quarter of gross return for both. The one-dollar commission floor per order binds at these order sizes, and accounts for twice as large a share of turnover at NOK 50,000 as at NOK 100,000. A strategy that looks profitable gross can therefore be unprofitable net purely because of portfolio size, without anything about the strategy having changed.

5.3 What the measurement itself taught us

The most transferable part of this work is not the results, but how many of them depended on the infrastructure rather than on the strategy.

On four occasions a test was green on a rule it did not exercise, as described in Chapter 3.4, and none of them would have been found by code review. A fifth error, that the portfolio engine skipped rebalancing dates falling on weekends, was not found by any test at all. It was found because the row count in a results file did not match the number of dates stated in the text. That error alone changed pure

momentum by 1.16 percentage points per year and reversed the sign of a conclusion about the ranking band.

The data would have given a different and better-looking answer had it been sourced more cheaply. The measurement in Chapter 3.2.1 shows that a free source is missing 45% of the index members from 2005, that almost all the missing names are delisted, and that the source shows the corrected accounting version in two out of three cases where a correction exists. Both errors pull in the same direction, and both would have looked like a better strategy. It is worth noting that neither produces an error message.

And biases that cannot be removed can at least be measured and given a direction. Survivorship through missing data coverage favours multifactor relative to pure momentum, that is, the very strategy the comparison is about, with a relative risk of 2.73 and p equal to 2.1 times 10 to the power of minus 7. The magnitude is at the same time small enough that the effect on the return figures is probably marginal. Both facts belong to the finding.

5.4 What the report cannot say

The absence of statistical significance is not proof that the strategies are alike. The excess return of multifactor over pure momentum has p equal to 0.96, and multifactor against the benchmark has p equal to 0.50. With a monthly standard deviation of 3% in the difference, far more than 326 observations are required to demonstrate a difference of a couple of percentage points. 27 years is a lot in practice and little statistically, and the sample cannot settle the question in either direction.

The entire volatility argument moreover rests on three stress episodes: 2000, 2008 and 2020. All the drawdown figures in the report are in reality three observations. Multifactor's advantage was in practice won in one of them. Without the dot-com episode there is not much left of the difference against pure momentum, and an argument resting on a single event cannot be strengthened by testing more parameters on that same event.

The factor regression described in Chapter 3.3 has not been run. It would be able to show whether what is left of the return in multifactor is exposure to known factors, which can be obtained more cheaply through index funds. It cannot change the headline result, since the strategy already loses to the benchmark, but it is missing.

Finally, the general caveats in Chapter 3.5 still apply. A backtest does not reflect latency, partial fills, market impact or discipline under pressure, and the fact that a specific implementation has no edge is not the same as the approach being wrong.

5.5 Implications for the other areas

The work on the systematic strategies passes three things on to the other areas.

The point-in-time contract and the portfolio engine are infrastructure that is not tied to the strategies tested here. Any strategy that ranks companies on accounting figures, that is, the core of fundamental

analysis, can be run through the same layer and thereby inherits no-look-ahead enforcement, historical index membership and explicit cost and tax modelling. That is the durable deliverable from this area.

The cost and tax figures transfer directly to risk management and portfolio construction. That friction takes around a quarter of gross return at these capital levels, and that tax is the single largest mechanism at low turnover, is a modelling assumption for any portfolio construction in the company, not a result that applies only here.

And the methodology for assessing a result transfers as well. A pre-specified set of criteria, a parameter sweep rather than a single point, a split into first and second halves, and an explicit assessment of whether a finding is large enough to be distinguished from noise. Those four steps uncovered more weaknesses in this report than the return figures themselves did.

6. Conclusion and the way forward

The research question was whether there are algorithmic trading strategies that, given realistic transaction costs and tax conditions for a Norwegian investment company, can deliver risk-adjusted returns exceeding those of a passive benchmark portfolio. The answer is no for the strategies tested here. None of the five families beat its benchmark on risk-adjusted returns after costs and tax.

Hypothesis 1, that filtered RSI-based mean reversion strategies can deliver positive risk-adjusted returns after transaction costs, is falsified for the conditions tested. The extension to four other strategy families did not change the conclusion, but made it more broadly supported.

6.1 What stands as findings

Four things stand as findings rather than as choices, that is, as results that hold across parameter choices and not only in the single run in which they were observed.

Sector neutralisation reduces risk substantially, and the fundamental blocks add little beyond that. Ranking within industry group accounts for 99% of the improvement in drawdown and 79% of the improvement in volatility between pure momentum and multifactor. Random groups reproduce only 19 and 30% of the two, so it is the actual industry classification that does the work. That 29 of 30 combinations in the parameter grid have lower volatility than pure momentum is a property of the construction, but it is the sector term in the construction that carries it.

More names is better than fewer, both for return and for declines. Concentration in fifteen names costs, and it costs most in volatility.

The momentum block contributes least in this universe over the full period. But it was the strongest block in the second half, so the claim applies to 1999 to 2026 as a whole and not to the future.

And selecting parameters on historical data points the wrong way. The correlation between the first and second halves is minus 0.72 for multifactor and plus 0.25 for factor rotation. In both cases a choice made halfway through would have produced a worse configuration than a random choice.

6.2 Recommendation

No single combination is promoted to production on this basis. Quality plus value with 25 names is the only configuration in the report that beats the benchmark, with 9.27% against 8.46% and a shallower drawdown, but it was selected after all thirty combinations had been run, in a grid where selection on historical data demonstrably points the wrong way. Picking it would be doing exactly what the report documents one should not do.

The passive part of the portfolio is therefore retained as it is, in EEA-domiciled UCITS funds under the participation exemption. This is not resignation. In this sample the benchmark has both the highest arithmetic return and the lowest volatility, that is, both at once, and an active strategy must beat it on both to justify the complexity and risk of operating it.

The decomposition nevertheless points to one measure that is cheap enough to consider on its own. Sector neutralisation requires no fundamental data, only an industry classification derived from a publicly reported code, and it delivers almost the entire reduction in declines. Should the company at some later point run a systematic equity strategy, diversification across industries is the first thing that should be in place, and it should be in place regardless of whether there is a fundamental dataset underneath. The converse also holds. A commercial dataset is by far the largest cost in the setup, and in this sample it pays for 0.65 percentage points per year, part of which is again a filter effect rather than signal.

6.3 The way forward

The next step should not be more variants on the same 27 years. The entire volatility argument rests on three stress episodes, and multifactor's advantage was won in one of them. More parameters tested on the same event provide no new information.

It should instead be an independent sample, that is, a different universe. For HULC AS this is moreover the relevant universe, namely European listings under the participation exemption. A multifactor sweep on an EEA universe would provide both a fresh out-of-sample test and a result that can actually be traded. If the patterns from Chapter 4.7 hold there, they are worth believing. If they do not, this grid is explained as a regime effect.

The decomposition also changes what should be tested first on the new universe. The most informative step is not another multifactor variant, but running the same ladder there, namely pure momentum, momentum with a sector constraint, and multifactor. If the split between the terms holds, sector diversification is a general finding and the fundamental data an expensive add-on. If it reverses, this sample is explained.

Three measurements remain before the picture is complete. The factor regression in Chapter 3.3 has not been run, and should be run on both single-stock strategies. The European commission disadvantage has not been measured, and the choice of market in Chapter 3.1.3 stands as an open methodological item until it is. And the NOK 300,000 threshold for reconsidering the choice of market should be replaced by a calculated threshold once both effects can be measured in the same simulation.

What carries forward in any case is the framework. The point-in-time contract, the portfolio engine with cost and tax modelling, and the validation methodology with mutation testing and pre-specified criteria are built to carry strategies other than those tested here. The clearest lesson of the report is that a backtest result is to a large extent a property of the measuring apparatus, and that an apparatus whose weaknesses one knows is worth more than a result one cannot verify.

7. Data and reproducibility

The single-stock strategies are run on S&P 500 members with historical index membership, 55 rebalancings from 30 June 1999 to 30 June 2026, with fundamental data in originally reported form filtered on availability date. The ETF strategies are run over the period January 2018 to December 2024, and the intraday strategies over January to June 2024. The cost model is set out in Chapter 3.1.2 and the tax treatment in Chapter 3.1.3. All parameter choices are stated in the section presenting the individual test. The framework is under version control, and every measurement has been run against a full test suite.

Our own measurement results are not attributed to sources, since they were produced in this work. The reference list below covers the claims the report borrows from others, that is, theory, prior empirical findings, definitions of key figures, commission rates and statutory provisions.

8. References

Journal articles

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.

Books

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 and software

Aroussi, R. (n.d.). yfinance [Python library]. https://github.com/ranaroussi/yfinance

French, K. R. (n.d.). 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. (n.d.). Commissions - stocks. https://www.interactivebrokers.com/en/pricing/commissions-stocks.php

Sharadar. (n.d.). Core US Equities Bundle (tables SF1, SP500, TICKERS and SEP) [dataset]. Nasdaq Data Link. https://data.nasdaq.com/databases/SFA

Legislation

Act on Taxation of Wealth and Income (the Norwegian Taxation Act), LOV-1999-03-26-14, Section 2-38. https://lovdata.no/lov/1999-03-26-14