Seasonality-Based Trading in Commodity Futures: A Risk-Aware Comparison of Dummy Variable Regression, SSA and RLSSA

Bachelor's Thesis

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Informatics

Author
Ralph Kosch
21-934-252
Ralphpeter.Kosch@uzh.ch

Professor
Prof. Dr. Anikó Hannák
Social Computing Group
Department of Informatics
University of Zurich

Supervisor
Robin Forsberg

Date of Submission: 30.11.2025

Abstract

This thesis investigates whether seasonal patterns in commodity futures can be turned into systematic trading strategies. Using monthly data on front-month contracts for 15 liquid commodity futures from 2001–2024, it evaluates three seasonality models in a backtest that incorporates realistic trading frictions. The results show that seasonality can generate economically meaningful excess returns in a few specific model–side combinations, while many configurations lose their edge once costs and risk are taken into account. The thesis concludes that commodity seasonality is exploitable only in carefully chosen setups and provides a Python framework for future strategy design and evaluation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Financial markets now operate at machine speed. Algorithms route and execute a large share of trades and quantitative funds rely on systematic strategies that scan data, extract patterns, and act with minimal human intervention. Reinforcement learning and deep reinforcement learning offer flexible frameworks that integrate prediction, signal generation, portfolio construction and execution within a single architecture (Sahu et al. 2023).

Alongside these more complex models, straightforward rule-based strategies continue to play a central role in practice. Momentum strategies, for instance, buy recent winners and sell recent losers over intermediate horizons, typically relying on six to nine months of past returns and rebalancing on a monthly basis (Song and Balvers 2022). Another important class, examined in greater depth in this thesis, is seasonality strategies. These exploit recurring calendar patterns by asking, for example, whether an asset that has historically performed well in a particular month is likely to outperform again in that same month (Rozeff and Kinney 1976). A well known example is the January effect, first documented in 1976 where average stock returns in January are abnormally high compared to the rest of the year (Rozeff and Kinney 1976).

The common thread is the use of transparent, testable rules that are grounded in data and executed mechanically. When designed and validated carefully, such rules can discipline discretionary judgment and may enhance risk-adjusted performance (Sahu et al. 2023). Turning historical regularities into reliable profits is difficult because markets evolve through regime shifts and time-varying risk premia, and implementation must handle transaction costs, liquidity constraints, and other frictions (Sahu et al. 2023).

1.2 Problem Statement

Building on this background, this thesis narrows the focus from general rule-based strategies to seasonality in commodity futures, which link financial markets to the real economy. Prices respond to macro conditions and investor flows, but they also encode harvest cycles, extraction schedules, inventories, logistics, and weather (Sørensen 2002). These real rhythms can generate repeating patterns that show up in calendar returns. The literature documents rich seasonal structure in commodity term structures and energy markets (Moreno et al. 2019). However, when tested transparently and consistently, well-known equity seasonals such as the Halloween or "Sell in May" effect (Magnusson 2020) do not give rise to a simple, persistent timing premium in broad commodity indices (Degenhardt and Auer 2018). This thesis therefore asks whether seasonality in commodity futures can still be used as a systematic signal to outperform a commodity benchmark on a classic and risk-adjusted basis.

Three approaches dominate in the literature for identifying time-series seasonality:

- **Dummy Variable Regression (DVR).** DVR models returns with an OLS regression that uses calendar dummy variables. The coefficients then represent the average seasonal premia and their statistical significance, which makes it possible to detect seasonality in the time series (Degenhardt and Auer 2018).

- **Singular Spectrum Analysis (SSA).** A data-driven decomposition that separates trend, seasonal components, and noise without a fixed parametric form. SSA can recover multi-frequency seasonality and improve forecasting accuracy in economic and commodity settings (Cipra 2020; Meng et al. 2024).

- **Robust $\mathbf{L_1}$ SSA (RLSSA)** This methods replace least-squares objectives in SSA with robust formulations to handle outliers and structural breaks, which can improve decomposition and forecast stability under contamination (Kazemi and Rodrigues 2023; Centofanti et al. 2025; Meng et al. 2024).

Research gap. For DVR, existing commodity seasonality evidence is limited to index-level tests of the Sell in May (SIM) effect rather than general monthly return patterns (Degenhardt and Auer 2018). For SSA and RLSSA, prior work emphasizes reducing forecasting error and documenting seasonal components rather than testing investable out-of-sample performance in commodity futures with explicit rolls, liquidity filters, turnover and costs (Kazemi and Rodrigues 2023; Centofanti et al. 2025). There is no unified comparison that (i) implements DVR, SSA, and RLSSA at the contract level across a broad commodity set, (ii) evaluates them against a transparent equal-weight benchmark, and (iii) adds an additional risk-aware evaluation layer that penalizes volatile commodity tickers. This thesis fills that gap by re-testing the three approaches with and without this risk-aware overlay in the current market environment to assess whether these models create value in realistic trading scenarios.

1.3 Research Objective

Goal. We test whether systematic, data-driven seasonality models can turn recurring structure in commodity futures into consistent excess returns under realistic constraints, and evaluate whether a risk-aware ranking overlay improves reliability and interpretability.

Hypotheses

1. **Hypothesis 1:** Seasonality strategies using DVR, SSA, and RLSSA deliver robust out-of-sample performance relative to a transparent equal-weight benchmark.
2. **Hypothesis 2:** A risk-adjusted specification that normalises monthly returns by contract-specific volatility produces more stable selections and portfolio paths for DVR, SSA, and RLSSA than the corresponding classic specifications.

1.4 Significance and Contribution

This thesis delivers a single, reproducible framework for detecting and evaluating seasonality under realistic trading conditions and shows how signals perform once they are forecast, traded, and compared with transparent benchmarks, transaction frictions, and strict out of sample validation. The architecture is built on open source data and operates at the contract level with explicit rolling rules and liquidity filters, which lowers the entry barrier for commodity futures research and allows other researchers to plug in new models and combinations without heavy data processing. This establishes a practical baseline for future studies and enables fast and comparable testing without bespoke pipelines. The findings also provide methodological guidance for academia and decision ready insight for institutional investors and risk managers who assess systematic strategies in commodity markets.

1.5 Scope and Limitations

The study focuses on monthly returns of front-month commodity futures with sufficient liquidity from 2001 to 2025. Daily and intraday horizons are excluded to target calendar-based seasonality and to reduce high-frequency noise. Portfolios follow Top 1 long and short selections without leverage or dynamic weighting. This design emphasizes clarity and reproducibility, but it may understate the potential of more complex allocation schemes. Throughout the analysis, transparency, robustness, and replicability take precedence over aggressive optimization.

Chapter 2

Theory

This chapter provides the theoretical foundation for the models and evaluation metrics used in the thesis. It first reviews key statistical properties of financial return series and explains why monthly data are suitable for identifying seasonality in commodity futures. It then introduces three methodological frameworks commonly used to detect seasonality signals: Dummy Variable Regression, Singular Spectrum Analysis, and Robust L_1 SSA. Finally, it presents the risk-aware evaluation layer, introducing the Sharpe ratio and the volatility modelling used to construct it. Together, these components define the analytical structure underlying the empirical results in the subsequent chapters.

2.1 Distribution Assumptions and Return Measurement

2.1.1 Empirical Rejection of Normality

In traditional financial theory, exemplified by early models such as the Option Pricing model by Black and Scholes 1973, it was common to assume that stock prices follow a geometric Brownian motion, which implies log-normally distributed prices and approximately normally distributed returns. However, empirical evidence now rejects the independent and identically distributed Gaussian view for financial return series. Engle and Patton 2001 show that asset returns are strongly non-Gaussian, with unconditional return distributions that exhibit heavy tails, non-zero skewness and clear departures from normality in standard tests, including kurtosis well above the Gaussian benchmark of three and skewness significantly different from zero.

Moreover, volatility is neither constant nor independent over time. Large shocks tend to be followed by periods of high volatility and small shocks by periods of low volatility, a pattern known as volatility clustering (Engle and Patton 2001). This empirical regularity motivates models with explicitly time-varying volatility such as ARCH/GARCH, EWMA and stochastic-volatility specifications (Engle and Patton 2001). As a result, models and inference procedures that rely on fixed-variance Gaussian assumptions can be seriously misleading. In line with these observations, this thesis focuses on modelling approaches that do not require restrictive distributional assumptions such as normality or log-normality, so that inference remains valid under the observed properties of financial returns.

2.1.2 Why Seasonality Analysis Uses Monthly Returns

In seasonal time-series modelling, a standard starting point is a pattern that repeats within a calendar year (Cipra 2020). In empirical same-calendar-month studies, returns are therefore measured at the monthly horizon and one tests whether average past returns in a given calendar month help to predict the current return in that same month (Keloharju et al. 2016; Y. Li et al. 2024). Although seasonal structure can occur at other frequencies, this thesis follows that established practice and works with monthly returns, which also helps to dampen high-frequency noise and microstructure effects that are more prominent in daily data.

2.1.3 Choice of Return Measure

For empirical implementation, monthly returns are often expressed in logarithmic form

$$r_{i,t}^{\log} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$$

where $P_{i,t}$ is the closing price of asset i in month t. An alternative is the simple return

$$r_{i,t}^{\text{simple}} = \frac{P_{i,t}}{P_{i,t-1}} - 1$$

The use of log returns here does not rely on any normality assumption. Even when normality is empirically rejected, log returns remain attractive because they are time-additive and therefore simplify multi-period aggregation, they respect the non-negativity of prices under compounding and they provide a natural interpretation as continuously compounded returns.

2.2 Seasonality Models

To detect seasonal patterns and make them operational for trading, this thesis employs three complementary frameworks: (i) time-series dummy-variable regression (DVR) to estimate month-specific return premia, (ii) Singular Spectrum Analysis (SSA) to extract data-driven seasonal components without imposing a fixed functional form, and (iii) a robust L_1 SSA (RLSSA) variant that reduces sensitivity to outliers and structural breaks. The next sections detail how each model is specified, how its signals are converted into concrete trading rules, and why it is suitable for detecting seasonality in commodity futures.

2.2.1 Time-Series Dummy-Variable Seasonality

Applicability

The time-series dummy-variable regression is a standard device for detecting calendar anomalies. It originated in equity markets but is now widely used across asset classes, including commodity futures. Its main advantage is interpretability: month-specific differences in average returns can be estimated with minimal structure. In essence, the regression tests whether the average return in a given month differs from the average in all other months.

A particularly relevant contribution for this thesis is Degenhardt and Auer 2018. Their paper revisits the “Sell in May” (SIM) effect, according to which returns tend to be higher in winter months (November–April) than in summer months (May–October). Unlike earlier studies that focused on broad equity indices, Degenhardt & Auer extend the analysis to include not only the Dow Jones Industrial Average which is an equity index but also commodity futures grouped in investable sub-indices of the Goldman Sachs Commodity Index.

Paper Methodology. The authors use dummy-variable regressions as the baseline specification and complement them with robustness checks based on Huber M-estimation and GARCH(1,1) to address outliers and time-varying volatility. They test the classic winter-versus-summer split and sector-specific versions, and they explicitly control for January effects that could confound SIM estimates.

Challenges. Several empirical issues make testing the SIM effect demanding. First, seasonal patterns can be distorted by the January effect or by influential outliers, which motivates robust estimation. Second, volatility clustering is common in financial returns, so GARCH models help verify that results are not driven by heteroskedasticity. Third, anomalies may weaken once widely publicized. The sample from 1989 to 2016 is therefore split into pre- and post-publication phases to evaluate persistence.

Findings. The study documents a statistically and economically significant SIM effect for DJIA stocks, robust across methods and strongest in industrials and basic materials. For commodity futures, the aggregate GSCI evidence is weaker, but some sub-indices show pronounced seasonality, in particular industrial metals. The authors also report subsample differences over time, consistent with changing market conditions and attention effects.

Relevance for this Thesis. This paper is central because it demonstrates that dummy-variable regressions can be applied to commodity futures, not only equities. By extending SIM beyond equities, Degenhardt & Auer show that systematic seasonal patterns in commodities can be detected with simple dummy structures, which supports the methodological choice made here.

DVR Methodology

The workflow has three steps. First, define a binary month indicator, also called a dummy variable. Set it to 1 when an observation falls in the target month and to 0 otherwise. This records whether the condition is true or false and includes the month explicitly in the regression.

Second, estimate a single OLS regression that includes this dummy. Although it is one model, the dummy splits the sample into the target month and all other months. The dummy coefficient is the difference between the average return in the target month and the average return in the remaining months. The intercept is the average outside the target month, and the intercept plus the dummy effect is the average in the target month. This is a difference in means framed as a regression.

Third, assess statistical significance with heteroskedasticity and autocorrelation consistent standard errors. For monthly data, a Newey–West estimator with a small lag is a common choice. This allows weak serial correlation and time varying volatility.

Equations

> *Our specification follows Degenhardt and Auer 2018 but we present it from intuition to implementation for clarity.*

Let r_t denote the asset's gross (or log) return in month t, for $t = 1, \dots, T$. For each calendar month $m \in \{1, \dots, 12\}$:

(i) Month indicator.

$$D_{t,m} = \begin{cases} 1, & \text{if month}(t) = m, \\ 0, & \text{otherwise.} \end{cases}$$

(ii) Two-group mean regression. Estimate by OLS:

$$r_t = \alpha_m + \beta_m \, D_{t,m} + \varepsilon_t, \quad t = 1, \dots, T.$$

Closed-form solutions are

$$\hat{\beta}_m = \bar{r}_m - \bar{r}_{\neg m}, \qquad \hat{\alpha}_m = \bar{r}_{\neg m},$$

where $\bar{r}_m$ and $\bar{r}_{\neg m}$ are the sample means in month m and in all other months.

(iii) HAC inference. Compute a Newey–West estimator of $\widehat{\ }(\hat{\beta}_m)$ with one lag and form

$$t_m = \frac{\hat{\beta}_m}{\sqrt{\widehat{\ }(\hat{\beta}_m)}}.$$

Under weak stationarity and finite fourth moments, $t_m \xrightarrow{d} N(0,1)$. Reject $H_0 : \beta_m = 0$ if $|t_m| > z_{1-\alpha/2}$.

Interpretation. $\hat{\alpha}_m = \bar{r}_{\neg m}$ is the average return outside month m. $\hat{\alpha}_m + \hat{\beta}_m = \bar{r}_m$ is the average return in month m. A significant $\hat{\beta}_m > 0$ means month m outperforms the rest on average. A significant $\hat{\beta}_m < 0$ means underperformance.

From estimates to a trading rule. Define the sets

$$\mathcal{M}^+ = \{m : \hat{\beta}_m > 0,\ |t_m| > z_{1-\alpha/2}\}, \qquad \mathcal{M}^- = \{m : \hat{\beta}_m < 0,\ |t_m| > z_{1-\alpha/2}\}.$$

Each year:

- Go long in the month $m \in \mathcal{M}^+$ with the largest $\hat{\beta}_m$.
- Go short in the month $m \in \mathcal{M}^-$ with the most negative $\hat{\beta}_m$.
- Stay flat otherwise.

Broader empirical context. Dummy-variable regressions for seasonality go back at least to Rozeff and Kinney 1976, who use monthly dummies to document the January effect in returns on the New York Stock Exchange. Since then, month dummies have become a standard tool to express calendar effects as differences in conditional means. The table below illustrates this development, from early equity evidence on the January effect to more recent methodological work on seasonal index estimation and normalization.

Table 2.1 – Empirical Studies Using Dummy-Variable Regression

Paper	**Empirical Methods**	**Problem Studied**	**Findings / Results**
Rozeff and Kinney 1976	Month–dummy style tests on NYSE. Full sample 1904–1974 with subperiods 1904–1928, 1929–1940, 1941–1974, and 1904–1928 & 1941–1974	Do monthly mean returns differ across the year?	Strong cross-month differences except in 1929–1940, driven primarily by high January returns. January dominates seasonal location shifts (e.g., Jan $\approx$ 3.5% vs. other months $\approx$ 0.5%). Implication: month dummies isolate a persistent January premium while cautioning about regime-specific attenuation.
(Cipra 2020)	Regression with $s-1$ seasonal dummies. Additive vs. multiplicative index normalization	How to estimate, normalize, and use seasonal components via regression?	Provides the regression recipe to obtain interpretable seasonal indices from dummy coefficients and to generate forecasts by extending dummies. Normalization rules: additive indices sum to 0. Multiplicative indices product to 1 (log-sum 0). Clarifies when indices are meaningful (regular, repeating seasonality) and how seasonality scales with trend in the multiplicative case.

Together, these studies demonstrate that the dummy-variable regression is a reliable tool for identifying seasonal return patterns. Its use in equity markets and its extension to commodity futures make it a natural and well-justified method for the empirical analysis in this thesis.

2.2.2 Singular Spectrum Analysis (SSA)

Applicability

Singular Spectrum Analysis (SSA) is a flexible, data-driven method for decomposing time series into interpretable components such as trend, oscillatory patterns, and noise. One of its basic tasks is to perform this decomposition with no a priori information about the time-series structure. In particular, SSA does not require an a-priori model for the trend or prior knowledge of the number and periods of seasonal components, which makes it especially useful when periodicities are uncertain or time-varying (Golyandina and Korobeynikov 2012).

A particularly relevant contribution for this thesis is Z. Li et al. 2024. They analyse weekly prices of 35 major commodity and precious-metal futures from five exchanges worldwide between January 2000 and September 2022 and use Singular Spectrum Analysis (SSA) as their main tool to uncover complex seasonal structure in futures prices and to study how this structure relates to market uncertainty.

Paper Methodology. Z. Li et al. 2024 first apply Basic SSA to the weekly price series of each commodity future. They construct the trajectory matrix, perform the singular value decomposition (SVD), and group eigentriples into trend, seasonal components, and noise. This decomposition yields several distinct cycles with different frequencies and amplitudes, including components with approximately quarterly, semi-annual, and annual periodicities. The presence and strength of these SSA-based seasonal components are then compared to the temporal aggregation levels used in their separate temporal hierarchical forecasting framework, which allows the authors to relate forecast accuracy at a given horizon to the degree of seasonality detected at the corresponding frequency.

Challenges. The paper emphasises that futures price series combine stochastic trends, high noise, and multiple overlapping seasonal components. Weekly futures prices can reflect harvest and storage cycles, production and inventory management, and calendar effects in trading, so relevant periodicities are not known in advance and need not be limited to one or two pre-set seasonal frequencies. Large exogenous shocks such as financial crises, pandemics, or geopolitical conflicts can also create changepoints and volatility spikes that temporarily dominate the regular seasonal pattern. Any seasonal model therefore has to distinguish relatively stable cycles from event-driven distortions without imposing an overly rigid parametric structure.

Intuition for SSA in Commodity Futures. Z. Li et al. 2024 argue that SSA is well suited to this setting because futures prices are noisy, non-stationary, and can embed several overlapping cycles whose frequencies are unknown ex ante. Production, inventory, and calendar effects can generate annual, semi-annual, quarterly, and longer cycles at the same time, so models that pre-specify only one or two seasonal frequencies risk missing relevant structure or forcing it into an inappropriate form. SSA instead lets the data determine how many periodic components exist and what their effective horizons are, while decomposing each price series into trend, multiple cycles, and residual noise in a non-parametric way.

Findings. The study documents that SSA identifies between three and nine cyclical or seasonal components in each futures series. Many of these components cluster around economically meaningful horizons, such as approximately quarterly, semi-annual, and annual periods. Tang et al. show that the amplitudes and shapes of these SSA-based seasonal components shift during major uncertainty events: for example, oil and nickel futures display stronger and more volatile seasonal cycles around the 2007–2009 financial crisis and the Russia–Ukraine conflict. They also report that temporal aggregation levels whose horizons match strong SSA seasonality tend to exhibit larger forecast accuracy gains in their temporal hierarchical framework, which suggests that SSA is capturing structure that is relevant for prediction.

Relevance for this Thesis. This paper is central because it applies SSA directly to commodity futures and suggests that strong SSA-based seasonal components are linked to better forecast performance at matching horizons. Z. Li et al. 2024 therefore provide a strong justification for using SSA as a core method to detect and analyse seasonality in commodity futures in this thesis.

SSA Methodology

The SSA workflow has four main steps, plus an optional forecasting step. First, the original series is embedded into a Hankel trajectory matrix that collects overlapping lagged vectors. Second, this matrix is decomposed via singular value decomposition, producing eigentriples that capture dominant patterns ordered by explained variance. Third, selected eigentriples are grouped to form low-rank approximations that correspond to trend, seasonal components, or noise. Fourth, each grouped matrix is converted back into a time series through diagonal averaging, yielding reconstructed components whose sum reproduces the original series.

To obtain forecasts, a fifth step extends the reconstructed series using a linear recurrence relation implied by the leading eigentriples. This recurrent SSA forecasting procedure produces multi-step-ahead predictions that respect the extracted data-driven structure.

Equations

> *Our specification follows Golyandina and Korobeynikov 2012 but we present it from intuition to implementation for clarity.*

Let $X_N = (x_1, \ldots, x_N)$ be a real time series of length N. Choose a window length L with $1 < L < N$ and set

$$K \;=\; N - L + 1.$$

(i) Embedding (trajectory matrix). Form the K lagged L-vectors

$$X_i = (x_i, \ldots, x_{i+L-1})^\top, \qquad i = 1, \ldots, K,$$

and stack them as columns to obtain the $L \times K$ trajectory matrix

$$\mathbf{X} \;=\; [X_1 : \cdots : X_K] \;=\; \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}.$$

This matrix is Hankel, meaning that all elements along each anti-diagonal are equal.

(ii) Decomposition (Basic SSA via SVD). In *Basic SSA*, we diagonalize the $L \times L$ matrix $\mathbf{X}\mathbf{X}^\top$ and use the resulting eigenpairs to construct a singular value decomposition of $\mathbf{X}$:

$$\mathbf{X} \;=\; \sum_{i=1}^{L} \sqrt{\lambda_i}\, U_i V_i^\top, \qquad \lambda_1 \geq \cdots \geq \lambda_L \geq 0.$$

Set $P_i = U_i$ and $Q_i = \sqrt{\lambda_i}\, V_i$, and define the rank-one matrices

$$\mathbf{X}_i \;=\; P_i Q_i^\top, \qquad i = 1, \ldots, L.$$

Then

$$\mathbf{X} \;=\; \sum_{i=1}^{L} \mathbf{X}_i, \qquad \|\mathbf{X}_i\|_F^2 \;=\; \|Q_i\|^2 \;=\; \lambda_i.$$

The triple $(\sqrt{\lambda_i},\, P_i,\, Q_i)$ is called the ith *eigentriple*.

(iii) Grouping and reconstruction. Let $d = \max\{j : \lambda_j \neq 0\}$. Partition $\{1, \ldots, d\}$ into disjoint index sets $I_1, \ldots, I_m$ and define

$$\mathbf{X}_I \;=\; \sum_{i \in I} \mathbf{X}_i, \qquad \mathbf{X} \;=\; \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}.$$

Grouping several leading eigentriples corresponds to an optimal low-rank approximation of $\mathbf{X}$ in the SVD sense and is typically used to capture trend and dominant seasonal components.

To map each grouped matrix back to a univariate series, apply *diagonal averaging* (Hankelization). For an $L \times K$ matrix $Y = (y_{lk})$, define

$$A_s = \{(l,k) : \, l + k = s + 1,\; 1 \leq l \leq L,\; 1 \leq k \leq K\}, \qquad s = 1, \ldots, N,$$

and set

$$\tilde{y}_s \;=\; \frac{1}{|A_s|} \sum_{(l,k) \in A_s} y_{lk}.$$

Applying this to $\mathbf{X}_{I_j}$ yields a reconstructed series $\tilde{X}^{(j)} = (\tilde{x}_1^{(j)}, \ldots, \tilde{x}_N^{(j)})$, and the original series decomposes as

$$x_n = \sum_{j=1}^{m} \tilde{x}_n^{(j)}, \quad n = 1, \ldots, N.$$

The $\tilde{X}^{(j)}$ are the elementary or grouped reconstructed components.

(iv) Recurrent forecasting. Fix a selected component or group defined by an index set I, and let $P_i \in \mathbb{R}^L$ be the corresponding eigenvectors. Write the first $L-1$ coordinates of P_i as P_i (abusing notation) and its last coordinate as π_i. Define

$$\nu^2 = \sum_{i \in I} \pi_i^2, \qquad R = (a_{L-1}, \ldots, a_1)^\top = \frac{1}{1-\nu^2} \sum_{i \in I} \pi_i \, P_i.$$

Define the extended series $Y_{N+M} = (y_1, \ldots, y_{N+M})$ by

$$y_i = \tilde{x}_i \quad (i = 1, \ldots, N), \qquad y_i = \sum_{j=1}^{L-1} a_j \, y_{i-j} \quad (i = N+1, \ldots, N+M),$$

so that $(y_{N+1}, \ldots, y_{N+M})$ is the M-step forecast based on the linear recurrence with coefficients a_j. In particular, for the one-step case, using the last $L-1$ reconstructed values from Step (iii),

$$\widehat{x}_{N+1} = \sum_{j=1}^{L-1} a_j \, \tilde{x}_{N+1-j} = a_1 \, \tilde{x}_N + a_2 \, \tilde{x}_{N-1} + \cdots + a_{L-1} \, \tilde{x}_{N-L+2}.$$

Broader empirical context. The previous subsection has shown how Z. Li et al. 2024 apply SSA to commodity futures, extract multiple seasonal components, and relate their strength to forecast performance at matching horizons. In this thesis we go one step further by using the standard recurrent SSA (SSA-R) forecasting (iv) step explicitly: for each futures contract we derive a linear recurrence relation (LRR) from the SSA decomposition, extend the reconstructed SSA signal via SSA-R, and translate the resulting seasonal pattern into trading rules. This design follows an SSA forecasting literature in which SSA-R is used to continue the signal and is evaluated directly against established time-series forecasting methods. Table 2.2 summarizes two representative studies that focus on SSA-based forecasting and report out-of-sample performance.

Table 2.2 – Empirical Studies on SSA-Based Forecasting Performance

Paper	**SSA forecasting setup**	**Forecasting task**	**Main SSA forecasting results**
Golyandina and Korobeynikov 2012	Basic SSA with reconstruction of signal components, followed by recurrent SSA (SSA-R) forecasting via linear recurrence relation (LRR).	Multi-step forecasts of real and simulated series with trend and multiple seasonalities (e.g. CO_2, MotorVehicle series).	SSA-R forecasts stay close to realised values for horizons up to 12 steps ahead when leading eigentriples are correctly grouped. Empirical guidance suggests $L \approx N/3$ as a good trade-off between signal capture and prediction error. Results indicate SSA-R is competitive with ARIMA-class models and robust in the presence of multiple periodicities.
Hassani 2021	SSA decomposition of the monthly US accidental deaths series, trend–seasonality reconstruction, and SSA-R forecasting for six future observations.	Six-month-ahead forecasting of the "Death series" (1973–1978), compared to SARIMA, ARAR, and Holt–Winters.	SSA achieves MAE = 180 and MRAE = 2%, outperforming SARIMA (MAE = 524, MRAE = 6%), Holt–Winters (MAE = 351, MRAE = 4%), and ARAR (MAE = 227, MRAE = 3%). Indicates SSA-based recurrent forecasting can reduce forecast errors by 2–3x in strongly seasonal settings.

Together these papers show that SSA-based recurrent forecasting is both methodologically standard and empirically effective, which supports our use of the SSA-R step to construct seasonality-based trading rules in commodity futures.

2.2.3 Robust L_1 Singular Spectrum Analysis (RLSSA)

Applicability

Robust L_1 singular spectrum analysis (RLSSA) extends classical SSA by replacing the standard least-squares singular value decomposition in the decomposition step with an L_1-based robust SVD. This makes the reconstruction and forecasts less sensitive to outlying observations and structural changes in the trajectory matrix, which can strongly distort classical SSA because the usual SVD relies on L_2-norm optimisation and allows a few extreme observations to pull the leading eigentriples towards themselves (Rodrigues et al. 2020; Kazemi and Rodrigues 2023). In contrast to Basic SSA, RLSSA targets the same low-rank seasonal and trend structure but downweights contaminated entries, so that a small fraction of unusual observations has limited influence on the recovered trend and seasonal components.

A particularly relevant contribution for this thesis is Rodrigues et al. 2020. Their study develops and applies an L_1-based robust SSA algorithm (RLSSA) and, as a companion approach, a Huber-based variant (RHSSA) to simulated series and to the daily quotas and returns of six large Brazilian mutual investment funds. The central message is that the RLSSA variant can recover signal and deliver accurate forecasts in the presence of outliers, while behaving similarly to classical SSA when contamination is small.

Paper Methodology. Rodrigues et al. 2020 first present the standard SSA pipeline of embedding, SVD, eigentriple grouping, and diagonal averaging. They then robustify the decomposition step by replacing the least-squares SVD with robust low-rank approximations: an L_1-based robust SVD that defines RLSSA and a Huber-loss SVD that defines RHSSA. For forecasting, they propose a three-stage robust recurrent SSA scheme. First, RLSSA (or RHSSA) is used to obtain a robust signal trajectory matrix by decomposing and reconstructing the series. Second, a standard SVD is applied to this robust signal matrix. Third, the leading left singular vectors are used to derive the linear recurrence relation (LRR) coefficients, which generate multi-step-ahead forecasts from the robustly reconstructed series. Forecast accuracy is evaluated using Root Mean Squared Error (RMSE) for both the mutual fund data and controlled simulation designs.

Challenges. Classical SSA implicitly assumes that deviations from the low-rank structure are roughly Gaussian and that large outliers are rare. Financial return series instead exhibit jumps, volatility bursts, and contamination from market shocks. Even a small fraction of extreme observations can seriously distort the least-squares SVD, produce spurious eigentriples, and bias the reconstructed components and forecasts. To study this effect systematically, Rodrigues et al. 2020 construct simulation scenarios with 2%, 5%, and 10% additive outliers and with 2%, 5%, and 10% multiplicative outliers, and compare classical SSA, RLSSA, RHSSA, and ARIMA across these contamination levels.

Intuition for RLSSA in Financial and Commodity Futures. RLSSA keeps the structural logic of SSA (trajectory matrix, low-rank decomposition, grouping, Hankelization, and LRR-based forecasting) but estimates the low-rank component with an L_1-based SVD. Under an L_1 objective, large residuals contribute linearly rather than quadratically, so isolated shocks and heavy-tailed innovations have much less influence on the estimated singular vectors. In settings such as financial and commodity futures, where seasonal and trend components coexist with abrupt price moves due to macro news, geopolitical events, inventory shocks, or liquidity squeezes, this is crucial. RLSSA is designed to recover relatively stable seasonal cycles while preventing a few extreme observations from dominating either the decomposition or the forecast.

Findings. When the data are essentially clean, SSA, ARIMA, and RLSSA achieve similar RMSEs for both model fit and forecasting, with SSA often slightly ahead for well-chosen window lengths. Once contamination is introduced, the picture changes: RLSSA and RHSSA clearly outperform classical SSA and ARIMA in both simulation and forecasting exercises. For example, in a synthetic series with 10% multiplicative outliers, RLSSA attains mean RMSE for model fit below 0.5, whereas classical SSA exceeds 3.0. In the mutual fund application, where contamination appears limited, RLSSA behaves similarly to SSA in tranquil periods but preserves stable forecast accuracy when large jumps are injected into the series in the simulation designs. The main trade-off is computational:

robust SVD is more expensive than classical SVD, though the authors report that RLSSA remains tractable for moderate window lengths that are typical in practice.

Relevance for this Thesis. For the purposes of this thesis, RLSSA is the key robust extension of SSA. It preserves SSA's ability to decompose futures prices into trend and multiple seasonal components and to construct LRR-based multi-step forecasts, while adding explicit protection against outliers and heavy tails. Commodity futures returns display extreme price moves, volatility clustering, and occasional regime shifts, so the robustness features documented by Rodrigues et al. 2020 provide a strong motivation to adopt RLSSA as the primary robust decomposition and forecasting tool for seasonality-based trading strategies in this thesis.

RLSSA Methodology

Conceptually, RLSSA follows the same four-step workflow as Basic SSA, plus an optional forecasting step. First, the original series is embedded into a Hankel trajectory matrix that collects overlapping lagged vectors. Second, instead of a classical SVD, a robust low-rank decomposition is computed, producing a signal matrix and a noise matrix. Third, the robust signal components are grouped and Hankelized to obtain a denoised reconstructed series. Fourth, a standard SVD is applied to the robust signal trajectory, and the leading singular vectors are used to derive a linear recurrence relation (LRR) that extends the fitted series and produces multi-step-ahead forecasts.

Equations

> *This section follows the approach of Kazemi and Rodrigues 2023. We keep the basic SSA notation: the trajectory matrix $\mathbf{X}$, the window L, $K = N - L + 1$, the index set I for selected components, and Hankel (anti–diagonal) averaging. The standard decomposition and forecasting steps are replaced by their robust counterparts.*

Let $X_N = (x_1, \ldots, x_N)$ be a real time series of length N. Choose a window length L with $1 < L < N$ and set

$$K = N - L + 1.$$

(i) Embedding (trajectory matrix). Form the $L \times K$ Hankel matrix

$$\mathbf{X} \;=\; [X_1 : \cdots : X_K] \;=\; \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix},$$

so $(\mathbf{X})_{lk} = x_{l+k-1}$ and all anti-diagonals are constant. This step is identical to Basic SSA.

(ii′) Robust decomposition (signal–noise split). Replace the classical SVD by a *robust* PCA/SVD. Denote the robust signal and noise trajectory matrices by $\mathbf{S}$ and $\mathbf{N}$, and write

$$\mathbf{X} \;=\; \mathbf{S} \;+\; \mathbf{N}.$$

Here, $\mathbf{S}$ is constructed from the robustly selected low-rank components (analogous to SSA's grouping), while $\mathbf{N}$ collects the remainder. In RLSSA, the low-rank approximation that defines $\mathbf{S}$ is obtained by minimizing an L_1-type loss rather than a least-squares loss, which limits the influence of outliers on the estimated singular vectors.

(iii) Grouping and diagonal averaging (Hankelization). Let I be the robust signal index set used to define

$$\mathbf{S} \;=\; \sum_{i \in I} \tilde{\mathbf{X}}_i,$$

so that $\mathbf{X} = \mathbf{S} + \mathbf{N}$. Reconstruct the denoised series by anti-diagonal averaging. For an $L \times K$ matrix $Y = (y_{lk})$ define, for $s = 1, \ldots, N$,

$$A_s = \{(l,k) : \, l + k = s + 1, \; 1 \le l \le L, \; 1 \le k \le K\}, \qquad \tilde{y}_s = \frac{1}{|A_s|} \sum_{(l,k) \in A_s} y_{lk}.$$

Applying this to $\mathbf{S}$ yields the robust fitted series $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_N)$ with $\tilde{x}_s = \tilde{y}_s$. Moreover, the Hankelized signal $\mathbf{S}$ and $\tilde{X}$ align column-wise:

$$\mathbf{S}_{:,k} \;=\; \big(\tilde{x}_k,\; \tilde{x}_{k+1},\; \ldots,\; \tilde{x}_{k+L-1}\big)^\top, \quad k = 1, \ldots, K.$$

In particular,

$$\mathbf{S}_{:,K} = \big(\tilde{x}_{N-L+1},\; \tilde{x}_{N-L+2},\; \ldots,\; \tilde{x}_N\big)^\top,$$

so the *tail* needed to start the recurrence is $(\tilde{x}_{N-L+2}, \ldots, \tilde{x}_N)$.

(iv′) Robust recurrent forecasting (three-stage RLSSA). The RLSSA forecasting step proceeds in three substeps:

1. **Robust reconstruction**

   Inputs from Step (iii) (kept fixed for forecasting): the robust signal matrix $\mathbf{S}$, the fitted series $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_N)$, and the column relation $\mathbf{S}_{:,k} = (\tilde{x}_k, \ldots, \tilde{x}_{k+L-1})^\top$ for $k = 1, \ldots, K$.

2. **Classical SVD on S (LRR coefficients)**

   Compute a standard SVD of $\mathbf{S}$ and retain r left singular vectors $U_1, \ldots, U_r \in \mathbb{R}^L$. For each j, let $U_j^{\triangle} \in \mathbb{R}^{L-1}$ be the first $L-1$ entries of U_j, and let $\phi_j \in \mathbb{R}$ be its last entry. Define

   $$\nu^2 \;=\; \sum_{j=1}^{r} \phi_j^2, \qquad \hat{a} \;=\; (\hat{a}_{L-1}, \ldots, \hat{a}_1)^\top \;=\; \frac{1}{1-\nu^2} \sum_{j=1}^{r} \phi_j \, U_j^{\triangle}.$$

   These coefficients define the $(L-1)$-term linear recurrence relation.

3. **Recurrent out-of-sample forecast**

   Initialize with the robust fitted series: $\hat{y}_t = \tilde{x}_t$ for $t = 1, \ldots, N$. Then for $h \geq 1$ generate

   $$\hat{y}_t \;=\; \sum_{j=1}^{L-1} \hat{a}_j \, \hat{y}_{t-j}, \qquad t = N+1, \ldots, N+h.$$

In particular, for the one-step case, using the tail from Step (iii),

$$\hat{y}_{N+1} = \sum_{j=1}^{L-1} \hat{a}_j \, \tilde{x}_{N+1-j} = \hat{a}_1 \, \tilde{x}_N + \hat{a}_2 \, \tilde{x}_{N-1} + \cdots + \hat{a}_{L-1} \, \tilde{x}_{N-L+2}.$$

Broader empirical context. The previous subsection has shown how RLSSA modifies SSA by introducing robust low-rank decomposition while retaining Hankelization and recurrent forecasting. Rodrigues et al. 2020 provide direct evidence that this robustification improves forecast performance under outliers and heavy-tailed noise. Subsequent work further extends and validates RLSSA in both univariate and multivariate settings. Table 2.3 summarizes two representative studies that deepen the methodological and empirical foundations of RLSSA.

Table 2.3 – Empirical and Methodological RLSSA Studies on Seasonality and Forecasting

Paper	**Empirical Methods**	**Problem Studied**	**Findings / Results**
Kazemi and Rodrigues 2023	RLSSA with iterative reweighted least squares and robust diagonal averaging	Robustification of SSA against outliers and jumps in financial time series	L_1-SVD recovers cycles more accurately under shocks. Simulations confirm superior outlier resistance relative to SSA. Seasonal components remain statistically significant at the 5% level.
Centofanti et al. 2025	Multivariate RLSSA (robust low-rank approximation applied diagonal-wise across series)	Extension of robust SSA to multivariate time series with heavy noise and cross-series dependence	Applying L_1 low-rank decomposition on marginals before aggregation improves seasonal extraction and yields robustness to heteroskedasticity and cross-series contamination.

Taken together, these studies show that RLSSA is both theoretically well founded and empirically effective. Its ability to preserve seasonal signals in the presence of outliers and structural breaks, combined with robust recurrent forecasting performance, makes RLSSA a natural and powerful tool for analysing commodity futures seasonality in this thesis.

2.3 Risk-Aware Evaluation

After describing the implementation of the three seasonality models, we now discuss how to evaluate seasonality-based trading strategies in a risk-aware way. Each model produces a one-step-ahead forecast of expected return based on seasonal patterns in the data. However, these forecasts scale with the volatility of the underlying contract.

2.3.1 Scaling Problem of Raw Seasonality Forecasts

To illustrate the scaling issue, we take the cocoa futures contract series as a representative example and use monthly simple returns from January 2005 to December 2014. From this return series $\{r_t\}$, we construct a scaled series by multiplying every observation by a constant factor c. In particular, we consider the original data with $c = 1$ and a rescaled version with $c = 2$, where all returns are doubled.

For each of the three models, we estimate the model on the original series and compute the corresponding statistics for both $c = 1$ and $c = 2$. For DVR the relevant quantity is the seasonal dummy coefficient $\hat{\beta}_m$. For SSA and RLSSA we report the one-step-ahead forecast. Table 2.4 summarises the results.

Table 2.4 – Effect of Scaling Monthly Returns by Factor c on Seasonality Models

	DVR $\hat{\beta}_m$	SSA forecast	RLSSA forecast
$c = 1$	0.0299894	−0.0192349	−0.0152709
$c = 2$	0.0599787	−0.0384698	−0.0305419

The pattern in Table 2.4 is the same for all three models. When the returns are doubled ($c = 2$), the DVR seasonal effect $\hat{\beta}_m$ and the SSA/RLSSA forecasts double as well. The underlying seasonal pattern, however, is unchanged and only the overall scale of the returns has been modified. At the same time, the risk of holding the contract has increased because the volatility of the underlying futures has also doubled. A ranking based purely on these raw outputs would therefore mechanically favour the more volatile version of the same contract, even though it does not contain a stronger seasonal signal.

Log returns instead of simple returns attenuate the effect somewhat, because the logarithm compresses large observations. Nevertheless, the forecasts from DVR, SSA and RLSSA still scale almost proportionally with the overall size of the returns. Log returns alone therefore do not solve the scale problem. Contracts with higher volatility still tend to generate larger raw forecasts and are mechanically preferred in a classic ranking.

2.3.2 Normalisation of Monthly Returns

To mitigate the scale problem at the input stage, we first normalise each contract's monthly return series before applying the seasonality models. This follows the SSA transformation literature. Hassani et al. 2020 show that, for monthly data and a one-step-ahead forecast horizon ($h = 1$), a standardisation step can improve SSA forecasting accuracy relative to using untransformed data. Motivated by this result, and by the need to make contracts comparable across different volatility levels, we standardise returns contract by contract.

Concretely, for contract i we define

$$y_{i,t} = \frac{r_{i,t} - \bar{r}_i}{s_i},$$

where $\bar{r}_i$ and s_i are the sample mean and standard deviation of contract i's returns over the estimation window. This transformation removes scale differences at the input stage. All contracts enter the seasonality models on a comparable unit-variance scale.

However, normalisation alone does not fully address the fact that different contracts have different levels of time-varying risk. For a fair, risk-aware comparison we also need an explicit model for conditional volatility, which we describe next.

2.3.3 EWMA Volatility Estimation

To obtain risk-adjusted decision scores, we need an estimate of each contract's time-varying volatility. As shown earlier (Subsection 2.1.1), returns exhibit volatility clustering, so using a constant variance would be inappropriate.

We estimate conditional volatility using an Exponentially Weighted Moving Average (EWMA) model following the RiskMetrics framework of J.P. Morgan 1996. For a return series $\{r_t\}$, the one-step-ahead variance forecast is updated as

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2,$$

with $0 < \lambda < 1$. A higher value of λ places more weight on older observations and produces smoother volatility estimates.

J.P. Morgan 1996 recommends $\lambda = 0.94$ for daily data and larger values for lower-frequency returns. For monthly data, their guideline is $\lambda = 0.97$. Following this recommendation, and to match the frequency of our commodity futures data, we set $\lambda = 0.97$.

The EWMA model is simple to implement, guarantees non-negative volatility estimates, and captures the persistent nature of volatility commonly observed in financial returns. The resulting monthly volatility forecasts serve as $\widehat{\sigma}_{i,t+1}$. In the portfolio construction stage, these $\widehat{\sigma}_{i,t+1}$ are combined with the de-standardised seasonality-based return forecasts to form risk-adjusted decision scores, so that a contract with twice the volatility must also have roughly twice the predicted return to be equally attractive. This removes the mechanical preference for high-volatility contracts and allows us to rank seasonality-based strategies in a genuinely risk-aware way.

Transition to methodology The theoretical toolkit is now complete. We have defined the return measures and introduced the three seasonality models together with the risk-aware extension. The next chapter turns to the implementation details and explains how this setup is used to generate the empirical results.

Chapter 3

Methodology

3.1 Data

3.1.1 Universe and Source

We study monthly log returns of front futures contracts across fifteen commodity markets from January 2001 to December 2024. Raw historical series were scraped from TradingCharts 2025, which lists individual monthly futures contracts per commodity. For each calendar month t and each commodity, we select the closest available futures contract whose average daily trading volume in that month exceeds a tradability cutoff of 1,000 contracts.

3.1.2 Tradability Filter

The tradability filter excludes contract–months that are so thinly traded that quoted prices may not reflect true market-clearing levels. Adämmer et al. 2016 argue that in commodity futures markets very low trading volume can hinder the incorporation of information into prices, so highly illiquid contracts may not reflect broader market conditions. At the same time, they show that in very thinly traded futures markets even some dozen transactions per week can suffice to generate reliable and informative price signals. Based on their cointegration and price discovery analysis they conclude that the trading volume threshold for efficient price discovery is very low and that even some dozen transactions per week may be sufficient to produce reasonably reliable price signals for production and storage decisions. Taken together, this suggests that reliable prices do not require extremely high volume, but that some minimum level of sustained trading activity is essential.

Against this background, we deliberately started from a conservative safety margin. In a first step, we only considered futures with an average of at least 500 contracts traded per day when constructing preliminary monthly series and checked whether this filter still allowed us to build continuous contract histories. We then gradually increased the cutoff to 1,000 contracts per day and found that no additional tickers were lost when moving from 500 to 1,000. Contracts that fail the final 1,000-contract criterion already lie well below the initial 500-contract threshold. We therefore adopt 1,000 contracts per day as our tradability cutoff. This level is far above the activity at which Adämmer et al. 2016 already report reliable price information, so it formalises a high but empirically grounded liquidity standard and ensures that included contracts are clearly active with plausibly tight bid–ask spreads.

3.1.3 Sample Coverage and Liquidity

Before presenting the return plots, Table 3.1 reports the observation window, the average monthly trading volume, and the minimum monthly volume for each market. Volumes are computed from the front most contract which accepts this threshold and is used to form each month's return.

In an earlier phase of this thesis we also considered palladium (PA) and platinum (PL) futures, but both contracts fell well below the 1,000-contract average daily volume threshold in the early part of the sample and therefore failed the tradability filter. As a result, PA and PL are excluded from the final universe reported in Table 3.1.

[0] The complete code base and processed datasets are available in the accompanying GitHub repository: `https://github.com/RPKosch/Seasonal-Trading-in-Commodity-Markets`.

Table 3.1 – Sample Volume by Future Contract (Jan 2001 – Apr 2025)

Ticker	Contract	Avg. Monthly Vol.	Min. Monthly Vol.
CC	Cocoa	13,575	2,857
CF	Coffee	15,248	4,063
CO	WTI Crude Oil	225,595	48,893
CP	Copper (High Grade)	33,177	1,006
CT	Cotton #2	13,834	1,061
GD	Gold	86,899	1,084
HE	Lean Hogs (Globex)	14,279	2,181
HO	Heating Oil	35,989	9,282
LE	Live Cattle (Globex)	18,686	3,969
NG	Natural Gas	71,478	9,241
SU	Sugar #11	48,129	8,101
SV	Silver 5000 oz	32,102	1,531
ZC	Corn (Globex)	117,352	11,093
ZS	Soybeans (Globex)	71,053	4,536
ZW	Wheat (Globex)	46,592	6,956

3.1.4 Return Plots for all Tickers

Figures 3.1, 3.2, and 3.3 show monthly long-hold return indices from 2001 to 2025, grouped by related markets.

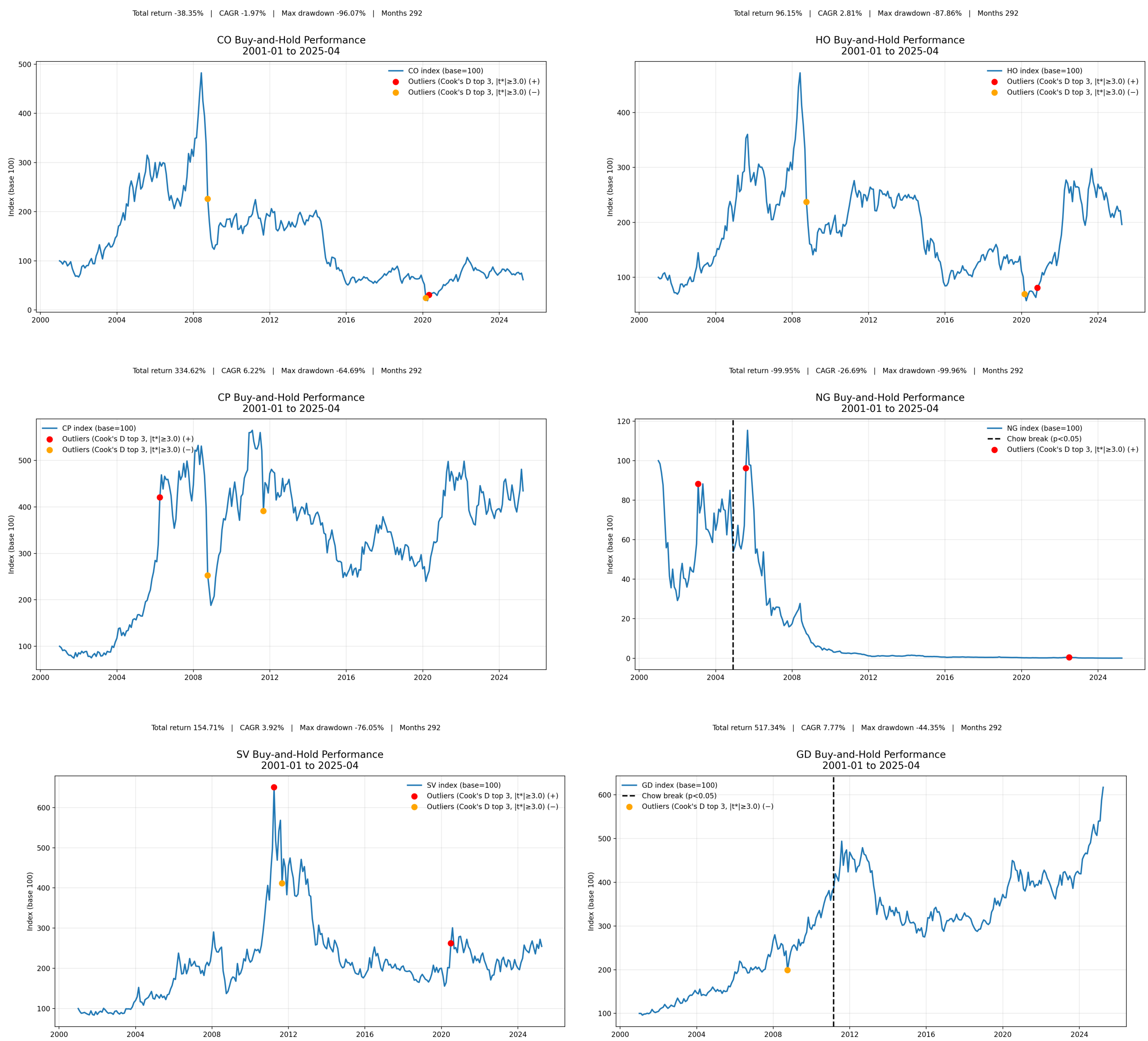

Figure 3.1 – Crude Oil (CO) and Heating Oil (HO) move almost one for one across the sample. Their long-hold indices rise and fall together over major energy cycles, consistent with both contracts being tied to the same physical supply chain and macro drivers. In contrast, Natural Gas (NG) is the clear outlier. The long-hold index shows repeated sharp spikes that are quickly unwound and, over the full window, loses almost all of its value. Irwin et al. 2020 document a very similar pattern for long-only natural gas ETFs. Despite average daily returns close to zero, high volatility, roll effects and product costs combine to produce a near-complete loss for buy-and-hold investors. Taken together, these patterns suggest that NG long positions tend to decay over time, even though the market still exhibits occasional, violent rallies that pose a risk to short positions.

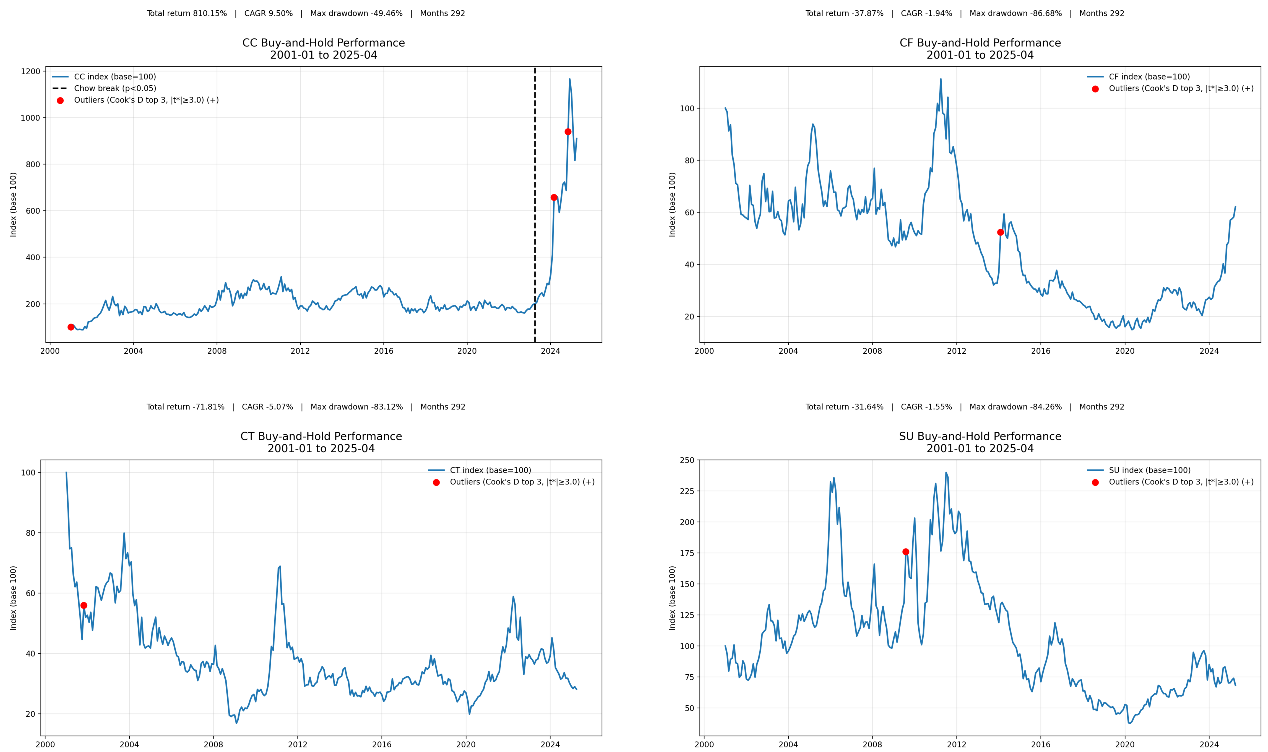

Figure 3.2 – Cocoa (CC) is the clear outlier among the soft commodities. For most of 2001–2023 the long-hold index fluctuates in a broad but bounded range. From late 2023 onward, however, the CC line breaks away and turns almost vertical. Futures prices explode upward and the long-hold index multiplies within a few months. Charry et al. 2025 document that international cacao prices more than tripled in 2024 as a sequence of severe supply deficits emerged in West Africa, driven by rising pest and disease pressure, extreme and poorly timed rainfall, and decades of underinvestment in ageing plantations, with speculative investor flows amplifying the move. Because cacao is produced by long-lived tree crops whose economic performance is highly sensitive to the initial establishment years, the global supply response is inherently slow, so prices can remain elevated rather than quickly reverting once the initial shock is absorbed. In the context of this thesis, the post-2023 cocoa episode illustrates how a multi-season supply shock and structural production constraints can dominate long-run buy-and-hold performance and overshadow gradual seasonal return patterns in an individual futures market.

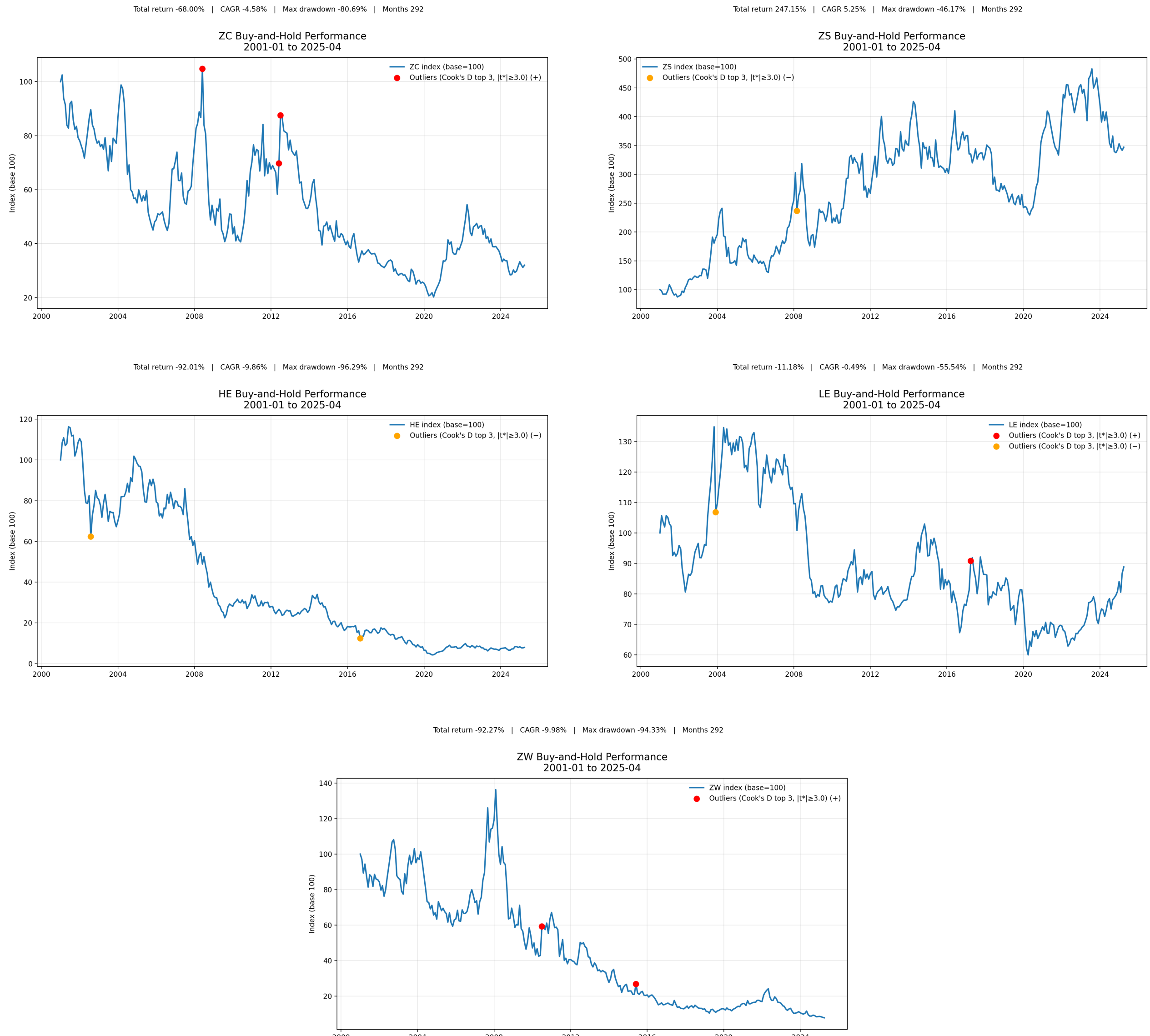

Figure 3.3 – Among the five contracts in this panel, two row crops stand out. Soybeans (ZS) show the strongest long-run appreciation: the long-hold index climbs from a base of 100 in 2001 to roughly four times that level by the mid-2020s. Martignone et al. 2024 describe soybeans as the most traded agricultural commodity worldwide and document sustained growth in global production from about 231 to nearly 390 million tonnes between 2008 and 2021, driven by expanding demand for protein feed, edible oils and biofuel inputs. In contrast, Chicago Wheat (ZW) drifts steadily downward after around 2009. Svanidze and urić 2021 show that the global wheat market moved into a very different regime after the 2007–08 and 2010–12 price spikes. From about 2013 onward, benchmark export prices trend lower even as world wheat exports almost double. They link this pattern to strongly supply-driven forces, namely a sequence of large harvests, yield improvements and the recultivation of formerly idle land, together with the rapid rise of Black Sea exporters such as Russia, Ukraine and Kazakhstan, whose low production and freight costs allow them to undercut traditional suppliers. The expansion and growing competitiveness of this region puts persistent downward pressure on international wheat prices, which is mirrored in the gradual erosion of the ZW long-hold index.

3.1.5 Why Diagnose the Data?

Before we can test for seasonality or evaluate trading rules, we need to understand the basic statistical properties of the underlying return series. If returns are nonstationary, exhibit strong volatility clustering, or contain structural breaks and extreme outliers, these features can bias inference and distort the performance of model-based signals. It highlights markets where additional caution is warranted and motivates specific modelling choices, such as the use of HAC standard errors.

The following subsections summarise five core diagnostics: stationarity of the mean process, volatility clustering, cross-market correlations, structural breaks, and return outliers. Together they provide a statistical backdrop for the seasonality analysis carried out in the remainder of the thesis.

3.1.6 Diagnostics: Stationarity

In this thesis we use stationarity in the applied sense of testing whether monthly log returns contain a unit root in their mean. If returns followed a random walk in the mean, the average level would drift over time and dummy–variable regressions would no longer measure stable calendar premia: month coefficients would partly absorb slow shifts in the mean, and the usual OLS t tests would lose their standard large–sample justification.

To assess mean stationarity we apply two complementary unit–root tests. The Augmented Dickey–Fuller (ADF) test treats the presence of a unit root as the null and rejects it when returns are sufficiently mean reverting. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test reverses the roles. Its null hypothesis is mean stationarity, and large, persistent deviations from a constant mean count as evidence in favour of a unit–root component. Together, ADF and KPSS indicate whether monthly returns are close enough to a stationary mean process for dummy–variable regressions to be interpreted as seasonality models. Remaining heteroskedasticity and short–run autocorrelation are handled with HAC standard errors.

Table A.1 reports ADF and KPSS p-values for all commodities. In every case the ADF test rejects the unit root at conventional levels and the KPSS test does not reject stationarity. Wee therefore can treat monthly log returns as stationary.

3.1.7 Diagnostics: Volatility

Beyond stationarity of the mean, it is essential to examine whether volatility itself is stable or subject to clustering. We use the ARCH LM test with 12 lags, which evaluates whether squared residuals are autocorrelated; rejection of the null indicates autoregressive conditional heteroskedasticity, i.e. volatility that changes systematically over time. The corresponding p-values for each commodity are reported in Table A.1 together with the stationarity diagnostics.

Overall, 8 out of 15 commodities reject the null of no ARCH effects at the 5% level, two cases are borderline, and the remaining 5 show no evidence of conditional heteroskedasticity. The series with clear heteroskedasticity are Cocoa, Crude Oil (WTI), Cotton, Natural Gas, Sugar #11, Silver, Corn, and Soybeans. Among these, Crude Oil, Cotton, Silver, and Corn in particular display very small p-values (below 0.002), indicating pronounced volatility clustering.

Because heteroskedasticity is present in several key commodities, it is necessary to account for it in the econometric analysis. In the empirical models of this thesis, inference is therefore based on Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) estimators.

3.1.8 Diagnostics: Cross-market Correlations

Understanding cross-market correlations helps to identify where diversification is available and where positions effectively load on the same underlying risk. Figure 3.4 reports the Pearson correlation matrix across all tickers. Two contracts stand out immediately: Natural Gas (NG) and Live Cattle (LE) are essentially uncorrelated with almost every other market, and even the correlation between NG and LE themselves is only mildly positive. This near-independence suggests that both contracts are driven by market-specific factors that are only weakly linked to the broader commodity universe.

Among the more connected assets, Crude Oil (CO) and Heating Oil (HO) show the strongest positive relationship, reflecting their joint role in the refined energy complex. Within agriculture, Corn (ZC), Soybeans (ZS), and Wheat (ZW) form a tightly knit cluster with relatively high mutual correlations. Precious metals also move together, with Silver (SV) and Gold (GD) displaying higher correlations than most other cross-commodity pairs.

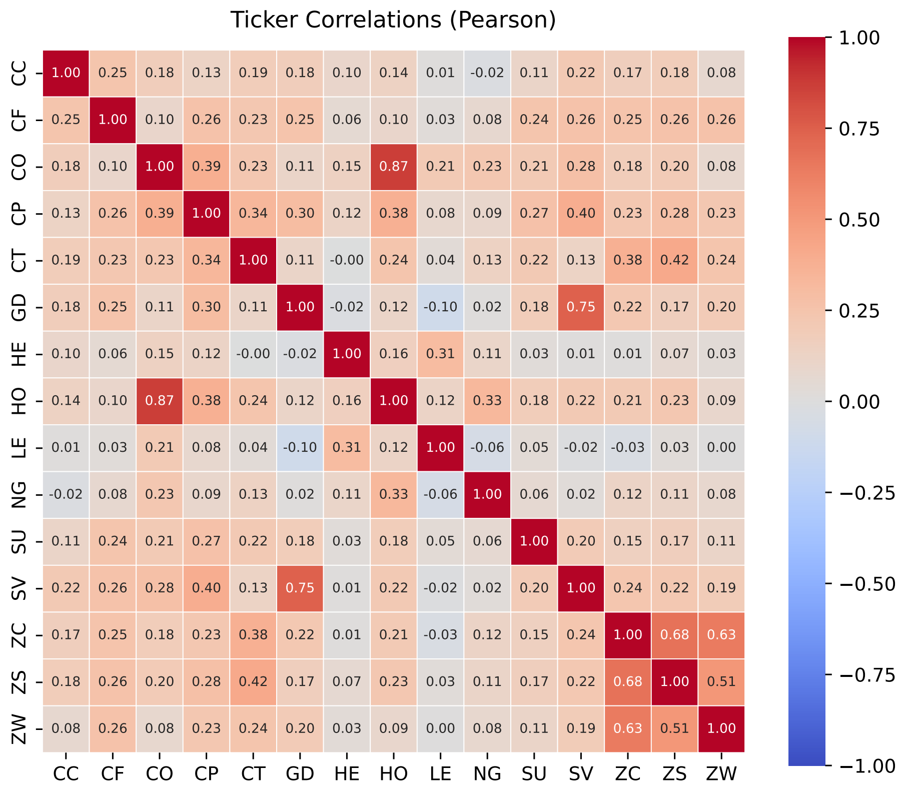

Figure 3.4 – Pearson correlation matrix of monthly returns (2001–2025).

3.1.9 Diagnostics: Structural Breaks

Structural breaks matter for seasonality because trading rules are calibrated on historical averages. If the mean return shifts after a major shock, pre-break seasonal patterns may no longer be informative for the new regime. We look for such shifts in the mean using a rolling Chow test for each commodity. For every admissible split date we compare one regression that uses the whole sample with another that allows different coefficients before and after that date, and we keep the split where this difference is most statistically significant. The last column of Table A.1 reports the strongest breakpoint and its p-value for each series.

As an example, the gold series exhibits a structural break around March 2011 that separates the crisis-driven boom from the subsequent normalization phase. Białkowski et al. 2013 show that from

2001 up to the all-time high in August 2011 gold is in a prolonged bull market largely linked to the global financial crisis (about 2007–2009) and the subsequent European sovereign-debt crisis (roughly 2010–2012). Bank failures, concerns about government solvency and very low interest rates make investors treat gold as a safe haven and strong inflows into gold-backed ETFs reinforce this demand, so rising uncertainty pushes gold's value steadily higher.

After the 2011 peak, Białkowski et al. 2013 document that their crisis and ETF-flow indicators stop intensifying and begin to level off as policy interventions gradually reduce the perceived risk of a full-scale meltdown. Our 2011 break is therefore naturally interpreted as a transition from a phase of growing crisis and expanding safe-haven premia to one of declining uncertainty, in which the exceptional demand for gold is slowly unwound.

3.1.10 Diagnostics: Outliers

Return outliers highlight extreme episodes where markets temporarily disconnect from their usual statistical patterns. These cases provide valuable insight into the economic forces behind tail risks. We detect them using Cook's distance, which measures how much the fitted regression would change if a single observation were removed. Months with a large Cook's distance are those that have an unusually strong influence on the estimated seasonal pattern. Table A.2 reports the strongest Cook's distance outlier for each commodity.

For crude oil (CO), the strongest regression outlier occurs in May 2020, when the front-month contract shows a monthly return of 61% and a Cook's distance of 0.125. This was the direct consequence of the historic WTI crash in April 2020, when prices briefly turned negative for the first time in history. Le et al. 2021 show that this crash was driven by a sudden collapse in oil demand due to Covid-19 lockdowns, while production remained high and a severe oversupply of crude oil quickly filled storage facilities, especially at Cushing. Large long investors in WTI futures who could not or did not want to take physical delivery were forced to close or roll their positions just before expiry, which pushed prices far below their normal range. The large positive return in May 2020 then reflects a rebound from these extreme levels as the most acute storage and rollover pressures passed and prices moved back toward more normal conditions.

3.2 Implementation of the Seasonal Strategy

With the diagnostic checks in place, we now turn to the implementation details. We first define the common data windows and parameter choices for DVR, SSA, and RLSSA, and explain how the resulting seasonality signals are translated into long and short portfolio positions. Next, we outline the two simulation setups so that we can test if our risk adjustment for the models improves performance. We then describe the construction of the equal-weight benchmark, and finally motivate the Monte Carlo layer used to quantify sampling variation and assess the statistical reliability of the performance comparisons.

3.2.1 Common data windows

Monthly log returns are used to estimate seasonal signals and monthly simple returns are used for compounding. Two important ranges are:

- **Lookback** (Rolling 10 years - Start 2006 – 2015): Used only to fit the seasonal signal for the next month.
- **Apply** (Start 2016 – 2024): Final out of sample evaluation.

3.2.2 DVR Variable Specification

In the DVR setup we use a two-sided Newey–West p-value with a 10% significance threshold. This choice is directly motivated by Degenhardt and Auer 2018, who study the "Sell in May" effect using dummy-variable regressions with Newey–West standard errors and classify the SIM effect as present when the relevant dummy has the expected sign and is statistically different from zero at least at the 10% level. Their empirical tables report 10%, 5%, and 1% significance stars for the SIM dummy, and the discussion treats $p \leq 0.10$ as sufficient evidence for a seasonal return anomaly in commodity index data.

3.2.3 SSA Variable Specification

In SSA we need to choose two tuning parameters: the window length L and the signal rank r. The window length L is the embedding dimension used to build the trajectory matrix and controls how many past observations enter each lagged vector. The signal rank r is the number of leading eigentriples that are treated as signal for reconstruction and recurrent SSA (SSA-R) forecasting.

For L we follow the guidance in Golyandina and Korobeynikov 2012. Their simulations and theoretical arguments show that forecasting works best when L is chosen smaller than $N/2$, and they highlight $L = N/3$ as a recommended value because it keeps forecast errors low while still capturing the main structure of the series. SSA theory, as summarised by Golyandina and Korobeynikov 2012 and Hassani 2021, further states that if a dominant seasonal period P is known, choosing a window length L that is divisible by or proportional to P improves the separability of the corresponding periodic component. In this spirit, for our monthly futures series with $N = 120$ and an annual period of $P = 12$ months, we set $L = 36 = 3P \approx N/3$, so each embedding window spans three complete seasonal cycles.

For the signal rank r we follow Hassani 2021, who also works with monthly data. As summarised in Table 2.2, SSA with $r = 12$ on the Death series delivers clearly better forecast accuracy than SARIMA, Holt–Winters, and ARAR. Since our monthly futures series with $N = 120$ has the same frequency, we adopt the same choice and set $r = 12$ to capture trend and several seasonal harmonics in a compact way. Z. Li et al. 2024 use much higher ranks (24 or 30 eigentriples) for very long weekly futures series with $N \approx 1{,}000$, but this reflects their richer multi-frequency weekly structure, so we regard the monthly evidence in Hassani 2021 as the more appropriate benchmark for our specification.

3.2.4 RLSSA Variable Specification

In line with Rodrigues et al. 2020, Kazemi and Rodrigues 2023 and Centofanti et al. 2025, we treat RLSSA as a drop-in robustified version of SSA and therefore do not retune the structural parameters. Concretely, RLSSA uses the same window length $L = 36$ and signal rank $r = 12$ as our SSA specification.

3.2.5 Classic versus risk-adjusted portfolios

For each seasonality model we take the specification above and consider two variants of the strategy. In the classic specification, DVR, SSA and RLSSA are estimated directly on monthly log returns, and the raw seasonality signal is used as the ranking score. In the risk-adjusted specification, we follow the risk-aware setup in Section 2.3. Before applying the models, we normalise each contract's monthly returns using

$$y_{i,t} = \frac{r_{i,t} - \bar{r}_{i,t}^{(10\text{y})}}{s_{i,t}},$$

as described in Subsection 2.3.2, where $\bar{r}_{i,t}^{(10\text{y})}$ is the rolling mean of contract i's monthly returns over our lookback period and $s_{i,t}$ is the current volatility level obtained from the EWMA model in Subsection 2.3.3. Both the classic and risk-adjusted variants are then traded using the same portfolio construction rules, which are explained in the next subsection.

3.2.6 Portfolio Construction and Trading Costs

Each month every model-based strategy, in both its classic and risk-adjusted version, holds one contract on the long side and one contract on the short side, namely the highest-ranked contract for the long portfolio and the lowest-ranked contract for the short portfolio.

Whenever we enter or exit a contract, we apply a proportional trading cost c. Following Marshall et al. 2012, who report that the average transaction cost for a \$1,000,000 commodity futures trade executed patiently over about one hour and only at the best quote ranges from 6.5 to 8.6 basis points (0.065%–0.086%), we deliberately take this upper bound and set $c = 0.00086$.

Given a monthly futures return r, the net long and short returns are

$$r_{\text{net}}^{\text{long}} = (1+r)\,(1-c)^2 - 1, \qquad r_{\text{net}}^{\text{short}} = (1+r)^{-1}(1-c)^2 - 1.$$

Portfolio values on both the long and short sides compound multiplicatively month by month using these net returns. Comparing the resulting classic and risk-adjusted portfolios then shows whether accounting explicitly for time-varying risk leads to an improvement in performance.

3.2.7 Benchmark

To judge whether the proposed seasonal strategies add value, we need a simple benchmark for reference. We therefore construct an equal-weight cross-sectional benchmark using all commodity tickers in our investment universe, applying the same monthly entry cost as for the model-based strategies.

This long-only series serves as the comparison portfolio and common reference for all evaluations. Short portfolios do not have a separate short benchmark and continue to use this long benchmark, which keeps all comparisons consistent and makes it clear whether the classic and risk-adjusted strategies outperform or underperform a simple equal-weight alternative.

3.3 Monte Carlo Layer

3.3.1 The Necessity of Monte Carlo Simulation

We want to measure how well the classic and these risk-aware seasonal strategies perform, both relative to each other and relative to the equal-weight benchmark. With a single realized historical path, each strategy produces only one full backtest, which does not allow us to quantify sampling variation or to assess statistical significance.

To overcome this limitation, we introduce a Monte Carlo simulation layer that, in each run, generates a completely new synthetic time series over the entire period for every contract. However, standard resampling or simulation techniques are not suited for commodity futures:

- Standard bootstrapping (resampling or shuffling returns) treats the data as independent and identically distributed (i.i.d.) draws. This breaks the serial dependence, seasonal behavior and local structure that the strategies are designed to exploit and therefore destroys the very autocorrelation patterns that drive signal stability (Vinod 2004).
- Gaussian simulation (adding $N(0,1)$ noise around an estimated mean) effectively assumes an i.i.d. Gaussian, zero-memory process, i.e. an $I(0)$ white-noise framework. Vinod and López-de-Lacalle 2009 argue that most economic time series are irreversible and non-stationary, making

such zero-memory $I(0)$ assumptions quite unrealistic. In our setting, simulating around a single unconditional mean would also flatten any seasonal pattern in expected returns and replace it by a constant mean plus white noise, thereby destroying the seasonal signals we want to study.

To address these issues, we employ the Maximum Entropy Bootstrap (MEB) algorithm of Vinod 2004 and Vinod and López-de-Lacalle 2009. MEB is a model-free resampling method designed for dependent, possibly non-stationary time series. Instead of fitting a parametric time-series model or forcing stationarity through detrending and differencing, it uses the Principle of Maximum Entropy to generate many plausible alternative versions of the observed series directly from the data. The resulting bootstrap paths reproduce the marginal distribution and basic time-dependence features of the original series such as autocorrelation and local peaks and troughs, while still being statistically distinct. This makes MEB well suited for our Monte Carlo layer, where we need realistic synthetic commodity-return histories that preserve the structure of the data without imposing a specific dynamic model.

3.3.2 The Maximum Entropy Bootstrap (MEB) Mechanism

Intuition The basic idea of the Maximum Entropy Bootstrap is simple:

- First, we temporarily forget about calendar time and just look at the values of the series, sorted from the smallest to the largest. From these sorted values we build a simple and smooth probability distribution which is still consistent with the data we have seen.
- Then we draw new values from this distribution and put them back into the original time positions according to their ranks. In other words, the days that were relatively high (or low) in the original series remain relatively high (or low) in the synthetic series, but the exact numbers are slightly different.

This produces new paths that share the same overall distribution and pattern of ups and downs as the original series, while still being genuinely new realizations.

Algorithmic Steps Formally, and following the description in Vinod and López-de-Lacalle 2009, we implement one MEB replicate of a time series x_t of length T in five steps:

1. **Separate values from time.** Sort the observations to obtain the order statistics

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(T)},$$

and store each value's time index. After this we know (i) which values occur and (ii) their time order.

2. **Build intervals between neighbouring values.** For each pair of adjacent order statistics $x_{(t)}$ and $x_{(t+1)}$ compute midpoints

$$z_t = \frac{x_{(t)} + x_{(t+1)}}{2}, \qquad t = 1, \ldots, T-1.$$

To extend the support, compute the absolute consecutive differences

$$d_t = |x_t - x_{t-1}|, \qquad t = 2, \ldots, T,$$

and let m_{trm} be a trimmed mean of $\{d_t\}$, discarding a small proportion of extreme absolute differences Vinod and López-de-Lacalle 2009 use 10% trimming in their toy example, i.e. trimming 10% of the observations in each tail). Set the lower and upper bounds to

$$z_0 = x_{(1)} - m_{\text{trm}}, \qquad z_T = x_{(T)} + m_{\text{trm}}.$$

Together, the points $z_0, \ldots, z_T$ define T non-overlapping intervals $[z_{t-1}, z_t)$, each associated with one sorted data point.

3. **Define a simple maximum-entropy distribution.** Assign probability mass $1/T$ to each of the T intervals and choose a simple density inside each interval, subject to a mean-preserving constraint (the average of the synthetic series must match the sample mean). This yields a piecewise-uniform density over the intervals $[z_{t-1}, z_t)$ with suitably chosen interval means.

4. **Draw new values in value-space.** Generate T Uniform$(0,1)$ random numbers u_t and sort them, $u_{(1)} \leq \ldots \leq u_{(T)}$. Associate the j-th sorted uniform $u_{(j)}$ with the j-th interval $[z_{j-1}, z_j)$, so each interval receives one point. Within that interval rescale to a local uniform variable

$$v_{(j)} = Tu_{(j)} - (j-1),$$

and map it linearly to a synthetic value

$$x^*_{(j)} = z_{j-1} + v_{(j)}\,(z_j - z_{j-1}), \qquad j = 1, \ldots, T.$$

This yields a new sorted sequence

$$x^*_{(1)} \leq \ldots \leq x^*_{(T)}$$

drawn from the fitted maximum-entropy density.

5. **Restore the original time order.** Finally, undo the sorting from Step 1: place each $x^*_{(t)}$ back into the time position of $x_{(t)}$. The resulting series $\{x^*_t\}_{t=1}^T$ is one MEB path. It has the same marginal distribution and preserves ranks of highs and lows, so peaks, troughs and seasonality are approximately retained.

MEB is defined for any sample size $T \geq 3$ (Vinod and López-de-Lacalle 2009). For shorter series, the intervals $[z_{t-1}, z_t)$ are wider and the synthetic paths can deviate more from the original. For longer series the intervals become narrower and each MEB replicate is a finer perturbation of the observed path. In both cases the bootstrap variation correctly reflects the amount of information in the data and provides an ensemble of realistic return histories for Monte Carlo inference.

Transition to results This concludes the design of the Monte Carlo layer. Together with the models, the ranking procedure, and the benchmark definition, it allows us to evaluate the strategies under a rich set of statistically plausible market evolutions. The next chapter presents the empirical results.

Chapter 4

Results

4.1 Results Overview

Before looking at concrete metrics and subperiods it is helpful to ask a very simple question. How often does a portfolio end the full 2016–2024 period with a gain. This win–loss view gives a quick sense of where each model is most reliable.

Table 4.1 – Win–loss counts by model and direction

Strategy	Long Pos	Long Neg	Short Pos	Short Neg
DVR Classic	100	0	100	0
DVR Risk-adjusted	100	0	100	0
SSA Classic	11	89	100	0
SSA Risk-adjusted	46	54	9	91
RLSSA Classic	19	81	72	28
RLSSA Risk-adjusted	10	90	45	55
Weighted benchmark	100	0	—	—

Table 4.1 shows three simple facts. Classic and risk adjusted DVR and the equal weight benchmark finish with a gain in every run so they are extremely robust. SSA and RLSSA are very asymmetric because their long portfolios are often negative while their short portfolios are almost always positive

in the classic specification. Finally the risk adjustment destroys the predictive power of SSA for the short series and turns a strong short signal into a weak and often loss making one.

4.1.1 Detailed Overview

With the basic win–loss picture in place we summarise full period statistics over 2016–2024 across 100 Monte Carlo runs. For each strategy and metric the tables report the average and median in one cell and the standard deviation of the run distribution in the next. We focus on cumulative return, the Sharpe ratio and maximum drawdown.

Table 4.2 – Full-period metrics across 100 runs, LONG side.

	CumRet		Sharpe		MaxDD	
Strategy	Avg / Median	SD	Avg / Median	SD	Avg / Median	SD
DVR Classic	0.2702 / 0.2268	0.1220	0.321 / 0.297	0.074	-0.336 / -0.342	0.026
DVR Risk-adjusted	0.1928 / 0.1897	0.0683	0.285 / 0.284	0.051	-0.402 / -0.406	0.031
SSA Classic	-0.3486 / -0.3833	0.2565	0.004 / 0.000	0.121	-0.719 / -0.713	0.064
SSA Risk-adjusted	0.0770 / -0.0406	0.4708	0.160 / 0.151	0.131	-0.721 / -0.724	0.048
RLSSA Classic	-0.2486 / -0.4261	0.6382	-0.027 / 0.008	0.258	-0.713 / -0.719	0.109
RLSSA Risk-adjusted	-0.4840 / -0.5999	0.4703	-0.168 / -0.161	0.273	-0.754 / -0.775	0.116
Weighted	0.1441 / 0.1444	0.0137	0.183 / 0.183	0.011	-0.415 / -0.415	0.003

On the long side the picture is simple. Both DVR portfolios beat the equal weight benchmark on return and Sharpe and they do this with a smaller maximum drawdown which is striking for such concentrated Top 1 strategies and points to a very robust signal. SSA and RLSSA do not deliver outperformance under our specification but the risk adjustment for SSA clearly helps and lifts its long portfolio to Sharpe levels that are close to the benchmark.

Table 4.3 – Full-period metrics across 100 runs, SHORT side.

	CumRet		Sharpe		MaxDD	
Strategy	Avg / Median	SD	Avg / Median	SD	Avg / Median	SD
DVR Classic	0.3439 / 0.2973	0.1693	0.328 / 0.319	0.059	-0.385 / -0.386	0.004
DVR Risk-adjusted	0.7244 / 0.7625	0.1863	0.444 / 0.465	0.056	-0.354 / -0.354	0.002
SSA Classic	2.6022 / 2.4663	1.6406	0.503 / 0.512	0.115	-0.584 / -0.591	0.065
SSA Risk-adjusted	-0.3087 / -0.3483	0.2751	0.067 / 0.090	0.125	-0.825 / -0.830	0.043
RLSSA Classic	1.4284 / 0.8519	2.4421	0.336 / 0.351	0.217	-0.738 / -0.761	0.130
RLSSA Risk-adjusted	0.0616 / -0.2207	0.8227	0.069 / 0.071	0.276	-0.702 / -0.706	0.121
Weighted	0.1441 / 0.1444	0.0137	0.183 / 0.183	0.011	-0.415 / -0.415	0.003

On the short side the ranking changes. We now have four portfolios that beat the benchmark on risk adjusted returns so our algorithms work much better on the short side than on the long side. The standouts are SSA classic and RLSSA classic which generate very large average returns but at the price of huge dispersion and very deep maximum drawdowns. Equally interesting is the risk adjusted DVR short. It earns about 72% over nine years with a relatively small standard deviation for such a high return and with milder drawdowns than the benchmark. Here the risk adjustment helps DVR a lot. For SSA and RLSSA the risk adjustment basically destroys their edge and leaves the short portfolios close to flat with weaker risk adjusted performance.

4.1.2 Subperiod results

We now examine how these patterns evolve over three non overlapping windows 2016–2018 2019–2021 and 2022–2024. We report subperiod means only and interpret averages in each window.

Table 4.4 – Subperiod means across 100 runs, LONG side.

		Classic			Risk-adj			Weighted		
Model	Window	CumRet	Sharpe	MaxDD	CumRet	Sharpe	MaxDD	CumRet	Sharpe	MaxDD
DVR	2016–2018	0.1207	0.411	-0.228	0.0341	0.212	-0.205	-0.1374	-0.542	-0.215
DVR	2019–2021	0.0177	0.189	-0.318	-0.2240	-0.604	-0.391	0.1855	0.452	-0.271
DVR	2022–2024	0.1139	0.363	-0.170	0.4865	1.005	-0.126	0.1189	0.361	-0.201
SSA	2016–2018	-0.6614	-1.426	-0.677	-0.6752	-1.385	-0.694	-0.1374	-0.542	-0.215
SSA	2019–2021	1.0689	0.960	-0.267	0.7173	0.779	-0.258	0.1855	0.452	-0.271
SSA	2022–2024	-0.0738	0.128	-0.628	0.9177	0.675	-0.391	0.1189	0.361	-0.201
RLSSA	2016–2018	-0.2804	-0.379	-0.460	-0.4459	-0.602	-0.553	-0.1374	-0.542	-0.215
RLSSA	2019–2021	0.2418	0.315	-0.394	0.1075	0.152	-0.366	0.1855	0.452	-0.271
RLSSA	2022–2024	-0.1601	-0.079	-0.642	-0.1686	-0.101	-0.601	0.1189	0.361	-0.201

Across subperiods the main message for longs is that Classic DVR is the only truly consistent winner. Classic DVR is the only strategy with a positive average return in all three windows which points to a robust seasonal pattern that survives very different market regimes. The benchmark and the risk adjusted DVR long only come out ahead in specific periods and do not match this consistency. In contrast SSA and RLSSA swing between deep losses and strong gains across the windows and this instability remains even after risk adjustment so with our specification they cannot deliver consistent return.

Table 4.5 – Subperiod means across 100 runs, SHORT side.

		Classic			Risk-adj			Weighted		
Model	Window	CumRet	Sharpe	MaxDD	CumRet	Sharpe	MaxDD	CumRet	Sharpe	MaxDD
DVR	2016–2018	-0.2960	-0.517	-0.385	-0.2007	-0.274	-0.354	-0.1374	-0.542	-0.215
DVR	2019–2021	1.1618	1.647	-0.141	1.1553	1.633	-0.072	0.1855	0.452	-0.271
DVR	2022–2024	-0.1180	0.067	-0.249	0.0004	0.197	-0.209	0.1189	0.361	-0.201
SSA	2016–2018	0.2830	0.442	-0.345	-0.2246	-0.381	-0.421	-0.1374	-0.542	-0.215
SSA	2019–2021	0.3833	0.420	-0.443	-0.2559	0.028	-0.734	0.1855	0.452	-0.271
SSA	2022–2024	1.0117	0.645	-0.371	0.2228	0.345	-0.403	0.1189	0.361	-0.201
RLSSA	2016–2018	0.0746	0.157	-0.380	0.0036	0.044	-0.381	-0.1374	-0.542	-0.215
RLSSA	2019–2021	0.0160	0.020	-0.574	-0.1797	-0.213	-0.548	0.1855	0.452	-0.271
RLSSA	2022–2024	1.2807	0.653	-0.365	0.3922	0.299	-0.310	0.1189	0.361	-0.201

Across short side windows the picture is different. Classic SSA and classic RLSSA have positive average returns in all three subperiods. The classic SSA algorithm is especially interesting because it shows very strong Sharpe ratios in every window and thus a high level of consistency for this strategy. Risk adjustment cuts returns and Sharpe for SSA and RLSSA in almost every window and does not reliably improve drawdowns so it removes most of their short side edge. Risk adjusted DVR short stands out as a more balanced alternative because it achieves solid returns with smaller downfalls and therefore scores better on risk adjusted evaluation.

4.1.3 Ticker Robustness

So far we have seen how the different portfolios perform. We also want to know how similar the classic and the risk adjusted versions are in terms of which contract they actually trade. To answer this we compute three simple summaries that are reported in Table 4.6.

Avg. C1% measures how concentrated the Top 1 choice is for the classic version. A value of 98% means that in a typical month ninety eight out of one hundred Monte Carlo runs end up with the same contract in first place. Avg. R1% does the same for the risk adjusted version. High values therefore mean that the model gives a very stable Top 1 signal while low values mean that small shocks often flip the winner. The Match rate tells us how often the classic and the risk adjusted version pick the same Top 1 contract in a month so it measures how much the risk adjustment changes the actual trade.

Table 4.6 – Top-1 signal stability by model and side: C1/R1 concentration and match rates.

Model	Side	Avg. C1 (%)	Avg. R1 (%)	Match rate (%)
DVR	Long	98.09	96.26	32.41
DVR	Short	99.07	98.66	39.81
SSA	Long	86.30	85.87	60.19
SSA	Short	85.44	86.73	55.56
RLSSA	Long	39.78	38.06	49.07
RLSSA	Short	42.41	38.51	43.52

Table 4.6 shows that the Top 1 signal is extremely stable for DVR and still very stable for SSA while RLSSA is much weaker. For RLSSA the much lower C1 and R1 values indicate that small changes in the underlying returns have a stronger effect on which ticker ends up in first place. It is also important to note that classic and risk adjusted versions have almost identical C1 and R1 values so the normalization layer does not really change how confident the model is in its first choice.

The Match rate between classic and risk adjusted versions is always around 50% which means that the two variants often behave in a similar way but still show clear differences in the chosen contract. Classic and risk adjusted DVR are especially interesting because they agree on the same Top 1 ticker in only about one third to two fifths of the months even though their aggregate returns are fairly similar. This suggests that the normalization layer often selects a different contract but still manages to achieve a very similar payoff pattern over time.

4.1.4 Significance

We now test which differences are statistically reliable. For each model and side we run paired two sided t tests across the one hundred runs at a significance level $\alpha = 0.01$. We compare the classic and risk adjusted versions and both against the equal weight benchmark.

Tables 4.7 and 4.8 summarise these tests. C denotes the classic specification, R the risk adjusted specification and W the equal weight benchmark. Entries show which portfolio is better together with the two sided p value, and n.s. marks cases that are not significant at $\alpha = 0.01$.

Table 4.7 – Paired tests on the LONG side with corresponding p-values.

	DVR			SSA			RLSSA		
Metric	Risk vs Classic	Wgt vs Classic	Wgt vs Risk	Risk vs Classic	Wgt vs Classic	Wgt vs Risk	Risk vs Classic	Wgt vs Classic	Wgt vs Risk
Return	C>R (0.0000)	C>W (0.0000)	R>W (0.0000)	R>C (0.0000)	W>C (0.0000)	n.s. (0.1551)	C>R (0.0034)	W>C (0.0000)	W>R (0.0000)
Sharpe	C>R (0.0000)	C>W (0.0000)	R>W (0.0000)	R>C (0.0000)	W>C (0.0000)	n.s. (0.0820)	C>R (0.0002)	W>C (0.0000)	W>R (0.0000)
MDD	C>R (0.0000)	C>W (0.0000)	R>W (0.0000)	n.s. (0.7286)	W>C (0.0000)	W>R (0.0000)	C>R (0.0077)	W>C (0.0000)	W>R (0.0000)

Long side. On the long side classic DVR stands out as the clear winner since it is significantly better than both risk adjusted DVR and the benchmark on all three metrics. Risk adjusted DVR still beats the benchmark so both DVR variants are statistically strong with the classic version being the best long only choice. For SSA the risk adjusted long is significantly better than the classic SSA on return and Sharpe but both SSA versions remain clearly weaker than the benchmark. RLSSA shows the opposite internal pattern because the classic version is significantly better than the risk adjusted one yet both RLSSA portfolios are still significantly below the benchmark on all three measures.

Table 4.8 – Paired tests on the SHORT side with corresponding p-values.

	DVR			SSA			RLSSA		
Metric	Risk vs Classic	Wgt vs Classic	Wgt vs Risk	Risk vs Classic	Wgt vs Classic	Wgt vs Risk	Risk vs Classic	Wgt vs Classic	Wgt vs Risk
Return	R>C (0.0000)	C>W (0.0000)	R>W (0.0000)	C>R (0.0000)	C>W (0.0000)	W>R (0.0000)	C>R (0.0000)	C>W (0.0000)	n.s. (0.3197)
Sharpe	R>C (0.0000)	C>W (0.0000)	R>W (0.0000)	C>R (0.0000)	C>W (0.0000)	W>R (0.0000)	C>R (0.0000)	C>W (0.0000)	W>R (0.0001)
MDD	R>C (0.0000)	C>W (0.0000)	R>W (0.0000)	C>R (0.0000)	W>C (0.0000)	W>R (0.0000)	n.s. (0.0528)	W>C (0.0000)	W>R (0.0000)

Short side. On the short side several portfolios are significantly stronger than the benchmark. Both classic and risk adjusted DVR beat the weighted portfolio on return, Sharpe and drawdown, with the risk adjusted DVR clearly the best within its group. Classic SSA and classic RLSSA also outperform the benchmark on return and Sharpe, but they do so with much deeper drawdowns. In contrast, risk adjusted SSA and risk adjusted RLSSA never beat the benchmark on any metric and are dominated by both their classic counterparts and the weighted portfolio. The benchmark has significantly smaller drawdowns than classic SSA and classic RLSSA, which shows that the very high short side premia in these strategies come with substantially higher volatility and tail risk.

Chapter 5

Discussion

This chapter interprets the empirical results in light of the two hypotheses stated in Section 1.3. The focus is on how well the three seasonality models (DVR, SSA, RLSSA) perform against the equal-weight benchmark and what we learn from the comparison between classic and risk-adjusted portfolios.

5.1 Revisiting the hypotheses

For convenience, the hypotheses are restated here:

- **Hypothesis 1:** Seasonality strategies using DVR, SSA, and RLSSA deliver robust out-of-sample performance relative to a transparent equal-weight benchmark.
- **Hypothesis 2:** A risk-adjusted specification that normalises monthly returns by contract-specific volatility produces more stable selections and portfolio paths for DVR, SSA, and RLSSA than the corresponding classic specifications.

5.2 Hypothesis 1: Seasonality-based strategies in commodity futures

5.2.1 What the results imply

From the results in Chapter 4 we see that, on the long side, the classic DVR Top-1 portfolio significantly outperforms the benchmark on a risk-adjusted basis, while SSA and RLSSA do not show a clear and consistent advantage. On the short side, by contrast, all three models (DVR, SSA, RLSSA), in both their classic and risk-adjusted specifications, outperform the benchmark on a risk-adjusted basis over the full sample.

We can therefore **accept** Hypothesis 1, with the important qualification that the empirical support comes mainly from the short side, whereas on the long side the result is essentially driven by the classic DVR specification.

5.2.2 Implications for DVR

Typical calendar effects that become widely known tend to lose their excess returns (Degenhardt and Auer 2018). In our case, using a broader DVR approach that considers all months rather than a single calendar window, we still obtain clear outperformance of the commodity benchmark over 2016–2024. Classic DVR long is profitable in all three subperiods with drawdowns similar to the weighted benchmark, which makes this simple specification an attractive long-side seasonality strategy.

5.2.3 Implications for SSA and RLSSA

Short-side performance. On the short side, SSA stands out as the strongest algorithm in our study. In all Monte Carlo runs the SSA short strategy ends the period with a positive outcome, and across all three subperiods it delivers high gains combined with very strong Sharpe ratios. This suggests that the SSA short implementation is robust in our commodity universe and provides a clear indication that SSA-based seasonality signals can be used in a practical investment environment on the downside.

Long-side underperformance and a possible explanation. On the long side, by contrast, both SSA and RLSSA perform clearly worse and do not manage to outperform the benchmark in a convincing way. A plausible hypothesis for this pattern is that the strategies often end up investing in relatively weak seasonal patterns simply because the rule is forced to hold a position every month. Even if SSA and RLSSA are able to extract seasonal structure, the Top 1 rule can still select contracts whose seasonal signal is small in absolute terms, which may dilute performance on the long side.

Possible improvements in design and parameter choice. This suggests natural directions for improvement. One idea is to introduce a pre-selection step based on signal strength. For example, a one-year pre-period could be used to measure the absolute seasonal amplitude for all long and short contracts. In the apply window, only contracts above a minimum absolute threshold would then be eligible for the Top 1 selection, such as the top 5% strongest absolute signals. This would avoid weak seasonality picks and might improve the SSA and RLSSA models.

If this still does not lead to a satisfactory improvement, a second idea is to allow for a variable selection of the rank parameter r for SSA and RLSSA. In the current implementation, r is a single, fixed choice for all contracts. The SSA literature indicates that performance is often sensitive to this parameter and proposes automatic rank–selection procedures (Kazemi and Rodrigues 2023). In applied work, Rodrigues et al. 2020 implement a data-driven choice of r by using w-correlation matrices to separate signal and noise components. Following this idea, a more flexible, contract-specific approach in our setting would estimate r separately for each series. Such tuning of r could help RLSSA in particular to realise more of its potential.

5.3 Hypothesis 2: Risk-adjusted specifications and stability

5.3.1 What the results imply

The evidence from our results does not support Hypothesis 2 as a general statement. The risk-adjusted specification improves performance only for DVR on the short side and for SSA on the long side. The better behaviour of SSA long after volatility normalisation is in line with Hassani et al. 2020, who report that standardisation enhances motnhly SSA forecasts with 1 step lookahead. For the short-side signals, however, our findings differ, as the risk adjustment does not lead to a clearer or more profitable SSA implementation.

For SSA and RLSSA overall, the risk-adjusted variants do not stabilise performance. On the contrary, they largely destroy the strong premia of the classic specifications on the short side, even though the SSA long portfolio improves in some metrics. We therefore **reject** Hypothesis 2 for our data set. It would still be interesting to retest this type of risk adjustment in combination with the improved SSA and RLSSA designs discussed above 5.2.3, since scaling returns by their volatility could in principle help to highlight clearer seasonal signals.

Chapter 6

Future Work

Building on the insights gained and the limitations uncovered in this thesis, several avenues for future research could deepen our understanding of seasonality in commodity markets and further enhance model performance. Paragraph 5.2.3 has already highlighted two concrete topics for improvement. The first is the optimization of the signal-strength filter that determines when a seasonal pattern is sufficiently strong to justify taking a position. The second is the contract-specific choice of the signal rank r, that is, the number of leading eigentriples treated as signal and thus the number of seasonal components used in the prediction step for each futures contract. Beyond these design refinements, there are several additional promising directions worth exploring.

6.1 Additional Tickers

This thesis focuses on 15 commodity futures tickers. Expanding this universe would increase the number of available signals for all models and could improve the Top 1 selection across specifications. Evidence from Adämmer et al. 2016 shows that in very thinly traded futures markets even some dozen transactions per week may suffice to generate reliable price information.

In the lookback period, it may therefore be useful to include additional tickers that are not yet liquid enough for actual trading with tight bid–ask spreads but still provide informative price data for modelling. This would allow us to exploit information from contracts that were less frequently traded in earlier years, provided their trading activity increases sufficiently before the start of our trading period.

6.2 Monthly Data vs. Daily Data

In this thesis the analysis is based on monthly data, which is a natural choice given the underlying theoretical framework. DVR should rather not be applied at the daily frequency due to the high level of noise, but SSA and RLSSA can be tested in a daily environment, as these methods explicitly separate signal from noise.

Future work can investigate whether SSA and RLSSA are able to identify clearer or more robust structures in commodity markets when applied to daily data with this additional information or whether monthly data remain superior in terms of model performance.

6.3 Portfolio Construction Beyond a Single Top-1 Contract

Throughout this thesis portfolios are constructed using a simple Top 1 rule. This provides a clear performance signal but leads to very concentrated portfolios. Future research could consider allocation rules that diversify across several contracts while maintaining a simple and transparent design.

Natural extensions include investing in the best N ranked contracts or in the top decile of signals, using equal weights or alternative weighting schemes. Combining such rules with the signal-strength filter discussed above would allow the strategy to trade only when signals are strong, while spreading risk across multiple tickers when positions are taken. This would help assess whether SSA, RLSSA, and DVR can deliver more stable excess returns once some diversification is introduced.

Chapter 7

Conclusion

7.1 Summary of Main Findings

The thesis investigated whether recurring seasonal patterns in commodity futures can be translated into investable classic and risk-aware trading strategies. Using contract-level data for fifteen liquid futures between 2001 and 2024, three models were compared: Dummy Variable Regression (DVR), Singular Spectrum Analysis (SSA), and Robust L1 SSA (RLSSA). Each model generated Top-1 long and short portfolios, which were benchmarked against an equal-weighted long-only commodity portfolio under explicit roll rules, tradability filters, and realistic trading costs.

The empirical results show that seasonal signals can generate economically meaningful and robust outperformance but only in specific model–side combinations. On the long side, the classic DVR specification delivers strong and persistent risk-adjusted outperformance of the benchmark, with stable contract selections and drawdowns of similar magnitude. On the short side, all three models perform more robustly but the classic SSA short portfolio clearly stands out in terms of stability and risk-adjusted returns, making it a particularly compelling basis for real investment implementation. In contrast, SSA- and RLSSA-based long portfolios are generally weaker and often fail to add value once costs are included.

Standardising returns through volatility scaling with an EWMA model helps in a few specific cases, such as DVR short and SSA long, but it does not yield a general improvement and can even weaken otherwise strong short-side signals.

7.2 Implications for Different User Groups

Researchers

There is still much to learn about seasonality and predictability in commodity futures. The empirical framework developed in this thesis, together with the accompanying Python repository, is fully reproducible and can serve as a solid base for further experiments and model development.

The results point to several concrete directions for future work with the proposed models. Most importantly, introducing an explicit signal-strength threshold would ensure that trades are executed only when seasonal effects are sufficiently strong. In addition, a data-driven procedure for selecting the SSA and RLSSA rank parameter r would allow the decomposition to adapt more flexibly to changing market conditions. Building on the existing code base, these extensions can be implemented and evaluated without altering the overall structure of the framework.

Institutional Investors

For institutional investors, the findings indicate that seasonality-based commodity strategies can be a useful and targeted addition to an existing portfolio. In particular an SSA-based strategy on the short side emerges as a strong candidate for implementation. It offers robust short-side premia and within a broader commodity allocation can contribute to diversification benefits because of its weak correlation with typical equity and long-only commodity exposures. Careful integration into existing risk and governance processes remains essential but the evidence suggests that such a component can add meaningful value.

Private Investors

For private investors and for those who use commodities mainly as a source of diversification, a easy approach is often best. Commodities should not be viewed as a primary driver of high returns but rather as an additional component within an equity-focused portfolio that can help reduce overall portfolio risk. This can be achieved by investing in a broad, equally weighted basket of commodities, for example through an index such as the SP GSCI Equal Weight Select Index, which is specifically designed to serve this diversification role.

7.3 Final Reflection

The main conclusion of this thesis is that seasonality in commodity futures is strong enough to be exploited in practice and in certain setups can realistically guide investment decisions. The most compelling evidence comes from trading the classic SSA model on the short side. More complex or seemingly more robust variations do not necessarily improve portfolio outcomes and can even dilute otherwise attractive signals.

The main contribution is therefore both methodological and practical: a reproducible, contract-level framework—supported by an open Python infrastructure—that allows researchers and practitioners to test seasonality models under realistic trading conditions and to see clearly where they add value and where they do not.

Within this framework, future work on signal thresholds, data-driven parameter choices, richer universes, and more diversified portfolio designs can build on the insights from Chapters 1–6 and continue to narrow the gap between statistical seasonality and implementable trading strategies.

Appendices

Appendix A

Tables

Table A.1 – Formal Diagnostics of Monthly Returns (2001–2025)

Commodity	ADF p-value	KPSS p-value	ARCH LM (12 lags) p-value	Strongest Break (date, p-value)
Cocoa	< 0.001 (supports)	$0.10 >$ (supports)	0.0119 (HE)	April 2023, $p = 0.0055$
Coffee	< 0.001 (supports)	$0.10 >$ (supports)	0.9945 (no ARCH)	no break
Crude Oil (WTI)	< 0.001 (supports)	$0.10 >$ (supports)	< 0.001 (HE)	no break
Copper	< 0.001 (supports)	$0.10 >$ (supports)	0.3075 (no ARCH)	no break
Cotton	< 0.001 (supports)	$0.10 >$ (supports)	0.0006 (HE)	no break
Gold	< 0.001 (supports)	$0.10 >$ (supports)	0.0617 (borderline)	Mar 2011, $p = 0.0014$
Lean Hogs	< 0.001 (supports)	$0.10 >$ (supports)	0.5036 (no ARCH)	no break
Heating Oil	< 0.001 (supports)	$0.10 >$ (supports)	0.0551 (borderline)	no break
Live Cattle	< 0.001 (supports)	$0.10 >$ (supports)	0.2809 (no ARCH)	no break
Natural Gas	< 0.001 (supports)	$0.10 >$ (supports)	0.0055 (HE)	Dec 2004, $p = 0.0407$
Sugar #11	< 0.001 (supports)	$0.10 >$ (supports)	0.0231 (HE)	no break
Silver	< 0.001 (supports)	$0.10 >$ (supports)	0.0011 (HE)	no break
Corn	< 0.001 (supports)	$0.10 >$ (supports)	0.0017 (HE)	no break
Soybeans	< 0.001 (supports)	$0.10 >$ (supports)	0.0444 (HE)	no break
Wheat	< 0.001 (supports)	$0.10 >$ (supports)	0.7509 (no ARCH)	no break

Table A.2 – Strongest regression outlier by Cook's Distance

Commodity	Date	Return	Stud. resid.	Leverage	Cook's D
CC	2024-03	0.595752	6.728697	0.040000	0.135742
CF	2014-02	0.419685	4.891646	0.040000	0.076796
CO	2020-05	0.610254	6.257168	0.041667	0.124845
CP	2008-10	-0.367347	-5.276211	0.041667	0.092041
CT	2001-11	0.252632	3.088828	0.041667	0.033545
GD	2008-10	-0.180978	-4.080038	0.041667	0.057122
HE	2016-09	-0.235217	-3.516988	0.041667	0.043067
HO	2020-03	-0.314933	-3.768024	0.040000	0.047079
LE	2003-12	-0.208131	-5.253319	0.041667	0.091316
NG	2003-02	0.522744	3.881225	0.040000	0.049804
SU	2009-08	0.303581	3.846902	0.041667	0.051100
SV	2011-04	0.305332	3.580899	0.040000	0.042720
ZC	2012-07	0.255253	3.891527	0.041667	0.052231
ZS	2008-03	-0.219524	-3.218935	0.040000	0.034814
ZW	2010-07	0.380282	4.551497	0.041667	0.070121

Bibliography

Adämmer, P., Martin T. Bohl, and C. Gross (2016). "Price Discovery in Thinly Traded Futures Markets: How Thin is Too Thin?" In: *Journal of Futures Markets* 36, pp. 851–869. DOI: 10.1002/fut.21760.

Białkowski, J., Martin T. Bohl, Patrick M. Stephan, and T. Wisniewski (2013). "The Gold Price in Times of Crisis". In: *Microeconomics: Production*. DOI: 10.2139/ssrn.1718106.

Black, F. and Myron S. Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81, pp. 637–654. DOI: 10.1086/260062.

Centofanti, Fabio, Mia Hubert, Biagio Palumbo, and Peter J. Rousseeuw (2025). "Multivariate Singular Spectrum Analysis by Robust Diagonalwise Low-Rank Approximation". In: *Journal of Computational and Graphical Statistics* 34.1, pp. 360–373. DOI: 10.1080/10618600.2024.2362222.

Charry, Andrés, Carolay Perea, Karen Ramírez, Guillermo Zambrano, Fredy Yovera, Adriana Santos, Tito Jiménez, Miguel Romero, Mark Lundy, Marcela Quintero, and Mirjam Pulleman (2025). "The bittersweet economics of different cacao production systems in Colombia, Ecuador and Peru". In: *Agricultural Systems*. DOI: 10.1016/j.agsy.2024.104235.

Cipra, T. (2020). "Seasonality and Periodicity". In: pp. 87–112. DOI: 10.1007/978-3-030-46347-2_4.

Degenhardt, Thomas and Benjamin R. Auer (2018). "The "Sell in May" effect: A review and new empirical evidence". In: *The North American Journal of Economics and Finance* 43, pp. 169–205.

ISSN: 1062-9408. DOI: 10.1016/j.najef.2017.09.003. URL: https://www.sciencedirect.com/science/article/pii/S106294081730089X.

Engle, R. and Andrew J. Patton (2001). "What good is a volatility model?" In: *Quantitative Finance* 1, pp. 237–245. DOI: 10.1088/1469-7688/1/2/305.

Golyandina, N. and A. Korobeynikov (2012). "Basic Singular Spectrum Analysis and forecasting with R". In: *Comput. Stat. Data Anal.* 71, pp. 934–954. DOI: 10.1016/j.csda.2013.04.009.

Hassani, Hossein (2021). "Singular Spectrum Analysis: Methodology and Comparison". In: *Journal of Data Science*. DOI: 10.6339/jds.2007.05(2).396.

Hassani, Hossein, Mohammad Reza Yeganegi, Atikur R. Khan, and E. Silva (2020). "The Effect of Data Transformation on Singular Spectrum Analysis for Forecasting". In: *Signals*. DOI: 10.3390/signals1010002.

Irwin, S., Dwight R. Sanders, Aaron Smith, and Scott Main (2020). "Returns to Investing in Commodity Futures: Separating the Wheat from the Chaff". In: *Applied Economic Perspectives and Policy* 42, pp. 583–610. DOI: 10.1002/aepp.13049.

J.P. Morgan (1996). *RiskMetrics™—Technical Document*. Tech. rep. New York: Morgan Guaranty Trust Company of New York. URL: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Kazemi, Mohammad and P. Rodrigues (2023). "Robust singular spectrum analysis: comparison between classical and robust approaches for model fit and forecasting". In: *Computational Statistics*, pp. 1–33. DOI: 10.1007/s00180-022-01322-4.

Keloharju, M., Juhani T. Linnainmaa, and Peter M. Nyberg (2016). "Return Seasonalities: Return Seasonalities". In: *Journal of Finance* 71, pp. 1557–1590. DOI: 10.1111/JOFI.12398.

Le, Thai-Ha, A. Le, and Ha-chi Le (2021). "The historic oil price fluctuation during the Covid-19 pandemic: What are the causes?" In: *Research in International Business and Finance* 58, pp. 101489–101489. DOI: 10.1016/j.ribaf.2021.101489.

Li, Yan, Qingfu Liu, Deyu Miao, and Yiuman Tse (2024). "Return Seasonality in Commodity Futures". In: *International Review of Economics & Finance*. URL: https://api.semanticscholar.org/CorpusID:269519598.

Li, Zhongfei, Wenjin Tang, Hui Bu, and Yuqiong Ji (2024). "Market uncertainty and information content in complex seasonality of prices". In: *Pacific-Basin Finance Journal*. DOI: 10.1016/j.pacfin.2024.102430.

Magnusson, Gylfi (2020). "Trick or treat? The Halloween effect in stock markets revisited". In: *Managerial Finance*. DOI: 10.1108/mf-12-2019-0622.

Marshall, Ben R., N. Nguyen, and Nuttawat Visaltanachoti (2012). "Commodity Liquidity Measurement and Transaction Costs". In: *Review of Financial Studies* 25, pp. 599–638. DOI: 10.1093/rfs/hhr075.

Martignone, G. B. Barboza, Bikramaditya Ghosh, Dimitrios Papadas, and Karl Behrendt (2024). "The rise of Soybean in international commodity markets: A quantile investigation". In: *Heliyon* 10. DOI: 10.1016/j.heliyon.2024.e34669.

Meng, Chen, Qingjie Du, and Haibing Shu (2024). "Return seasonalities in the Chinese stock market". In: *Pacific-Basin Finance Journal*. DOI: 10.1016/j.pacfin.2024.102391.

Moreno, Manuel, Alfonso Novales, and Federico Platania (2019). "Long-term swings and seasonality in energy markets". In: *European Journal of Operational Research* 279.3, pp. 1011–1023. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2019.05.042. URL: https://www.sciencedirect.com/science/article/pii/S0377221719304722.

Rodrigues, P., J. S. Pimentel, Patrick Messala, and Mohammad Kazemi (2020). "The Decomposition and Forecasting of Mutual Investment Funds Using Singular Spectrum Analysis". In: *Entropy* 22. DOI: 10.3390/e22010083.

Rozeff, Michael S. and William R. Kinney (1976). "Capital market seasonality: The case of stock returns". In: *Journal of Financial Economics* 3, pp. 379–402. DOI: 10.1016/0304-405X(76)90028-3.

Sahu, Santosh Kumar, Anil Mokhade, and Neeraj Dhanraj Bokde (2023). "An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges". In: *Applied Sciences* 13.3. ISSN: 2076-3417. DOI: 10.3390/app13031956. URL: https://www.mdpi.com/2076-3417/13/3/1956.

Song, Jian and Ronald J. Balvers (2022). "Seasonality and Momentum across National Equity Markets". In: *The North American Journal of Economics and Finance*. URL: https://api.semanticscholar.org/CorpusID:248874494.

Sørensen, Carsten (2002). "Modeling seasonality in agricultural commodity futures". In: *Journal of Futures Markets* 22, pp. 393–426. DOI: 10.1002/FUT.10017.

Svanidze, Miranda and Ivan urić (2021). "Global Wheat Market Dynamics: What Is the Role of the EU and the Black Sea Wheat Exporters?" In: *Agriculture* 11.8. ISSN: 2077-0472. DOI: 10.3390/agriculture11080799. URL: https://www.mdpi.com/2077-0472/11/8/799.

TradingCharts (2025). *Futures Historical Data*. https://futures.tradingcharts.com/. Accessed: 2025-09-09.

Vinod, H. (2004). "Ranking mutual funds using unconventional utility theory and stochastic dominance". In: *Journal of Empirical Finance* 11, pp. 353–377. DOI: 10.1016/j.jempfin.2003.06.002.

Vinod, H. and Javier López-de-Lacalle (2009). "Maximum Entropy Bootstrap for Time Series: The meboot R Package". In: *Journal of Statistical Software* 29, pp. 1–19. DOI: 10.18637/jss.v029.i05.

Statement on the Use of AI Tools

In preparing this thesis, I used AI-based language assistance tools exclusively for restyling sentences and for correcting grammar and spelling. All ideas, analyses, and conclusions are my own and I remain fully responsible for the content of this work.

Eidesstattliche Erklärung

Der/Die Verfasser/in erklärt an Eides statt, dass er/sie die vorliegende Arbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als die angegebenen Hilfsmittel angefertigt hat. Die aus fremden Quellen (einschliesslich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

....	....
Ort, Datum	Unterschrift des/der Verfassers/in