A Professional Journal Published by the International Federation of Technical Analysts

IFTA Journal 27

Inside This Issue

2 Trend Scoring - Assigning an Overall Technical Score to any Instrument Across Multiple Time Frames

14 Prediction is Very Difficult: Especially if it's About the Financial Markets! (Part III)

32 The Reliability of MACD and RSI Crossovers During Geopolitical and Economic Crises Events: A Multi-Asset Technical Study

56 Signal and Anomaly Detection in Noisy Time Series: A Comparative Study Using Financial, Energy, and Astronomical Data

70 The Role of Sentiment Analysis Using Artificial Intelligence and Market Cycles in Enhancing TASI Tactical Trading

94 Stationary, But Not Profitable? A Critical Look at Pairs Trading

113 Book Review: Predicting the Price

115 In Memoriam



"You can cut all the flowers but you cannot keep Spring from coming."—

—Pablo Neruda

ISSN 2409-0271

Letter From the Editor

By Dr. Rolf Wetzer, CFTe, MFTA



Dear Colleagues

A year has passed, and another issue of the IFTA Journal has been published. It seems to be a constant in a rapidly changing world. Just a year ago, the situation was marked by optimism and the vision of a positive future. That vision has begun to crack, and at the same time, hope is becoming an increasingly valuable asset.

Wars are dominating the media headlines. In addition to the conflict in Ukraine, another completely unnecessary war has been thrust into the spotlight since the beginning of the year. Since the U.S. attack on Iran, energy prices have been skyrocketing worldwide, causing global inflation spikes. Last year, it also became more difficult to distinguish between friend and foe. Relationships cultivated over many years are being cast aside by Stone Age-style economic sanctions. And to top it all off, in areas where human intelligence should be paramount, we're witnessing the triumph of artificial intelligence. We need hope now more than we have in a long time.

In times like these, it's good to know that with IFTA we have a platform where friendships are nurtured, experiences are shared across borders, and relationships are strengthened. People have been at the forefront for over 40 years now.

For the past 38 years, the best place to experience IFTA has been the annual conference, which will take place in London this year. The global technical analysis community returns to London for the 39th Annual IFTA Conference — a two-day gathering of the world's most influential thinkers in technical analysis, quantitative trading, and financial markets.

I would like to thank everyone who made this journal possible. First and foremost, of course, are the authors. Again, we have a colorful mix of articles from colleagues, MFTA papers, an educational section from Italy, and our book review from Australia.

Next, we need to thank NAAIM and Susan Truesdale. By now it has almost become a tradition that we are allowed to publish their papers. This year we have the pleasure to feature Davide Pandini, the winner of the NAAIM Founders Award. He is a member of the IFTA board and the author of our series on time series forecasting.

A major contribution comes from Linda Bernetich's production team. Year after year, they produce a publication with quality that never fails to impress. Finally, I would like to thank Regina Meani and Mohamed El Saiid for their time and effort in our team.

Best Regards,

Dr. Rolf Wetzer, CFTe, MFTA

IFTA Journal

EDITORIAL

Dr. Rolf Wetzer, CFTe, MFTA
Editor and Chair of the Editorial Committee
Rolf.Wetzer@ifta.org

Mohamed El Saiid, MFTA, CFTe
melsaiid@hc-si.com

Regina Meani, CFTe, MSTA, AMT
regina.meani@gmail.com

Send your queries about advertising information and rates to admin@ifta.org

Letter From the Editor

By Dr. Rolf Wetzer, CFTe, MFTA ..... Inside front cover

Articles

Trend Scoring - Assigning an Overall Technical Score to any Instrument Across Multiple Time Frames

By David Linton, MSTA, MFTA 2

Prediction is Very Difficult: Especially if it's About the Financial Markets! (Part III)

By Davide Pandini, PhD, MFTA, CFTe, CMT, CSTA 14

MFTA Papers

The Reliability of MACD and RSI Crossovers During Geopolitical and Economic Crises Events: A Multi-Asset Technical Study

Sarah Alyasiri, MFTA, CFTe 32

Signal and Anomaly Detection in Noisy Time Series: A Comparative Study Using Financial, Energy, and Astronomical Data

Patrick Plog, MScFE, MBA, MFTA, CFTe 56

The Role of Sentiment Analysis Using Artificial Intelligence and Market Cycles in Enhancing TASI Tactical Trading

Maha Saeed, MFTA, CETA, CFTe, MSC 70

NAAIM Paper

Stationary, But Not Profitable? A Critical Look at Pairs Trading

By Davide Pandini, PhD, MFTA, CFTe, CMT, CSTA 94

Book Review

Predicting the Price *by David Linton MFTA*

Reviewed by Regina Meani, CFTe, MSTA, AMT 113

Author Profiles 114

In Memoriam 115

IFTA Staff 117

IFTA Board of Directors 117



The IFTA Journal is published annually by the International Federation of Technical Analysts, 1300 Piccard Dr., Suite LL 14 Rockville, MD 20850 USA. © 2027 International Federation of Technical Analysts. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying for public or private use, or by any information storage or retrieval system, without prior permission of the publisher.

Trend Scoring - Assigning an Overall Technical Score to any Instrument Across Multiple Time Frames

David Linton, MFTA, MSTA
London, United Kingdom
david@updata.co.uk

Abstract

One of the most common areas of confusion within teams of investment professionals and communication with clients arises from not being able to clearly define trend and time horizon. This paper sets out a technical analysis framework whereby multiple time frames can be taken into account for evaluating the overall strength of trend for any instrument. It also seeks to define mechanical entry and exit rules for positions based on the scoring mechanism outlined.

The purpose of this trading strategy is to reduce confusion and ambiguity around different time horizons and ensure all investment participants, and potentially their clients, are on the same page in understanding overall trend state of an instrument. The scoring is unambiguous.

The paper outlines a traditional Dow Theory approach around multiple trends and understanding bullish, bearish and sideways trending markets. A clear framework is provided for defining long-term, medium-term and short-term trends and using this framework as a common language between people is strongly advocated.

A number of potential technical analysis techniques are proposed allowing the market professional to devise their own scoring mechanism. The methodology chosen here, as an example, utilises Ichimoku cloud charts for their relative simplicity in defining bullish, bearish and sideways trend states.

The soring mechanism assigns more significance to the longer-term primary trend and decreasing importance to the medium-term and shorter-term trends, though changes in these trend states are instrumental in making decisions to enter and exit positions. The objective is to understand when all the trends are lining up ahead of major market participation and when they are beginning to come to an end.

The proposed overall trend scoring mechanism allows for large universes on instruments to be ranked allowing the market professional to home in on outperformance quickly and easily. This same mechanism should ensure that a fund is holding the strongest trending instruments. This simple trend following strategy ensures 'the trend is your friend.'

Dow Theory — The founding rules of technical analysis

Over 100 years ago, Charles Dow, who founded the Wall Street Journal and devised the Dow Jones Index, identified that market price behavior followed trends. Six basic rules, which we now know as Dow Theory, came out of his work. These rules still apply in today's markets:

1. The market has three movements — primary, secondary and short-term trends
2. Trends have three phases — accumulation, participation and distribution
3. The stock market discounts all news
4. Stock market averages must confirm each other — divergence
5. Trends are confirmed by volume
6. Trends exist until definitive signals prove that they have ended otherwise

The ability to resolve the presence of multiple trends for an instrument is a skill that continues to allude many market participants. Prices may rise before they fall or fall before they rise. Timing the multiple trends in train can be the difference between success and failure. While the trend is your friend, all trends come to an end. There will be times where the short-term trend changes are indicating the potential for the longer-term trends to change as well and times where they are just a pullback in price. It is valuable to have a mechanism that can measure the totality of the multiple trends that are present.

Time Horizon

Defining whether a chart is bullish or bearish is not always that straight forward. Figure 1 shows a daily chart of American Express. Over the course of several months the price has been in an uptrend, but in the last two months the price has been trending lower. This highlights the fact that there are two trends in train here, a primary trend and a secondary counter trend. This emphasizes the importance of time horizon.

Figure 1 - Daily chart of American Express — Chart: Updata Analytics, Data: Bloomberg



One of the biggest advantages of technical analysis versus techniques such as fundamental analysis is the ability to change the chart time frame in order to change the time horizon of a view. A daily chart allows us to mentally extrapolate price action for the coming weeks. But this chart will not be very helpful for predicting what will happen in the next hour and it will probably be too detailed for making a prediction of where prices will be years out. Figure 2 shows how we can relate chart frame to time horizon.

Figure 2 — Time Horizon Table — Source: Cloud Charts — Trading Success with the Ichimoku Technique, 2010

	ULTRA SHORT	VERY SHORT	SHORT TERM	MEDIUM TERM	LONG TERM	VERY LONG	ULTRA LONG
Chart Interval*	Tick-1 min	5-10 mins	Hourly	Daily	Weekly	Monthly	Quarterly
Predict Forward	Minutes	Hours	Days	Weeks	Months	Years	Many Years
Lookback	Hours	Days	Weeks	Months	Years	Many Years	Decades

SHORT TERM TRADER LONG TERM INVESTOR	SHORT TERM	MEDIUM	LONG TERM	VERY LONG SHORT TERM	MEDIUM	LONG TERM	VERY LONG

The table sets out a framework whereby a common language is achieved between market professionals, trading and investment teams and their clients. A daily chart can be considered medium term with a time horizon of several weeks ahead. For this chart, it will normally be necessary to look back over several months, at least, in order to visualize where the prices are likely to be headed in the coming weeks. Thinking longer term, to make a judgement on where prices might be in the coming months, a weekly chart time frame will normally be better. In this case looking back at the chart data over a few years, at least, will be necessary. For predicting years ahead, a monthly chart is likely more suitable. Conversely, shorter term views require the analyst to look at intra-day chart intervals. A 60-minute chart time frame, short term, might contain just a few weeks of data and is more appropriate for predicting price movement more accurately for the coming days. A day trader, thinking hours ahead may use a 5-minute or 10-minute chart.

Different market participants will often have differing views of what short term and long term actually mean. For a long term buy and hold investor, short term may be a period of several weeks, while a short-term trader may consider a week to be long term. This highlights the pitfalls of using language such as long-term and short-term without defining the time horizon. Defining a view as 'The long-term weekly Chart, looking months ahead' provides clarity to anyone that a view is being communicated to. This also ensures that the right time frame chart is being analyzed for the desired time horizon. The easiest way to remember this is: minutes means hours, hours mean days, days means weeks, weeks means months and months means years. A quick recoil guide is shown in Figure 3.

Figure 3 — Chart Frame and Time Horizon — Quick Reference

VIEW	Minutes	Hours	Days	Weeks	Months	Years	Decades
means	⇧	⇧	⇧	⇧	⇧	⇧	⇧
CHART	Seconds	Minutes	Hours	Days	Weeks	Months	Years

Given the complexity of these multiple time frames, it is not normally accurate to simply say that an instrument is bullish or bearish. It will very often be the case that an instrument is bullish and bearish simultaneously on differing time frame charts. The only time that it can be said that an instrument is unequivocally bullish is if it is making an all-time high on all time frames. A historical low price point may be unanimously bearish. Just stating that an instrument is bullish or bearish without consideration of multiple trends risks being wrong on at least one time horizon and leaves room for misinterpretation.

Trend Identification

There are a number of technical analysis tools for defining the trend of a chart. Some are quite subjective, such as the process of drawing trend lines on a chart manually. Channels, such as least squared regression trends, require beginning and end points of the chart data to be chosen. Moving averages are one of the best-known tools, whereby prices above a moving average can be considered bullish and below the average is bearish. The subjectivity of these trend identification makes it difficult to feed values into a scoring model.

A second problem with techniques such as trend lines and moving averages for trend definition is the binary interpretation of a chart being bullish or bearish. Markets are not always trending up or down and it is often the case that prices become rangebound in a sideways direction. In this case prices will oscillate more rapidly above or below their moving average. This points to a sideways trend state between bullish and bearish. This phase may signal a change from bullish to bearish or vice versa or it may be a consolidation phase before the previous trend resumes. The shorter-term trend can provide clues as to which scenario is most likely, again highlighting the importance of having an overall understanding of the different time frame charts.

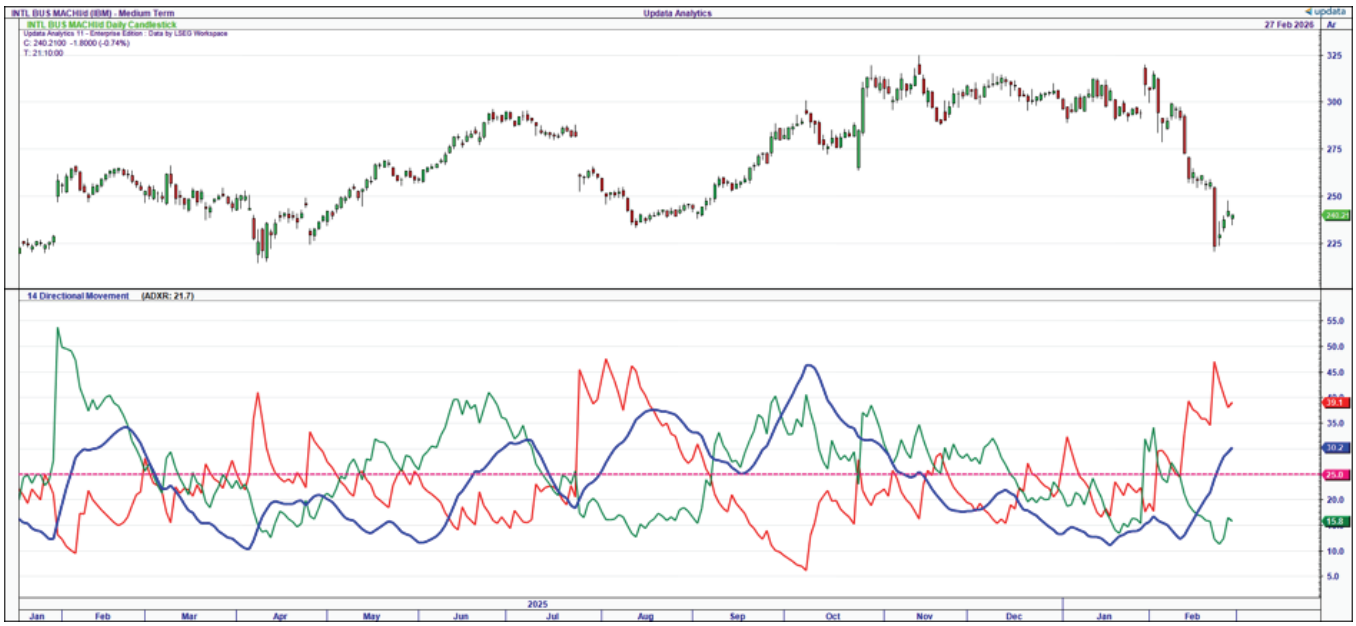
Bollinger Bands are a more dynamic way of seeing uptrends, downtrends and sideways trends as shown in the medium-term daily chart of IBM in Figure 4. If prices are between the average and the upper band, prices are trending up. Price between the average and the lower band, signals a downtrend. When prices oscillate between the bands with the bands themselves running horizontal, this can indicate a sideways market. Price rising or falling outside the bands indicates a dramatic increase in volatility. The bands coming together, what Bollinger calls 'the squeeze', normally occurs ahead of a volatility increase, and often heralds a price breakout or a change in trend. Figure 4, shows how the uptrends, downtrends and sideways trend typically last several weeks to a few months. Days means weeks.

Figure 4 — IBM Medium Term with Bollinger Bands — Chart: Updata Analytics, Data: LSEG



The technical analysis indicator, which seeks to objectively define if a market is trending or non-trending, sideways, is Welles Wilder's Directional Movement introduced in his groundbreaking 1978 book, New Concepts in Technical Trading Systems. Figure 5 shows a Directional Movement chart for IBM. Wilder advocated using a 14-day period, but it may make more sense to use a longer period such as 21 days to approximate a trading month. Most people quote the ADX value as defining the trend although Wilder's actual measure, which he termed ADXR, was the average of the current value for ADX and the ADX 14 periods ago.

Figure 5 — IBM Medium Term Directional Movement — Chart: Updata Analytics, Data: LSEG



The Directional Movement rules for defining the trend are as follows:

Rising ADX(R)

- Between 15 - 25 Beginning of Trending, use trend indicators
- Between 25 – 45 Definite Trending – trend indicators
- Over 45 Trend over extended – price or indicator patterns

Falling ADX(R)

- Below 20 No trend, use oscillators
- Between 20 - 30 Consolidation – use oscillators
- Between 30 - 45 Correction likely – price or indicator patterns

Whilst these rules are potentially of value, the main limitation with using this technique for trend definition is that changing the variable, will change the results in defining the trend. Wilder's 14-days, which is the default setting on most systems, may be too short a period for reading the trends on most daily charts. Many analysts using Directional Movement fail to appreciate this and just accept the default setting of 14-days. Furthermore, do these settings carry well to different time frame charts. Directional Movement does effectively signal a change between bullish to bearish trends with the +DI and -DI lines crossing each other, but again getting the period to match the data effectively will be key. Most technical analysts use Directional Movement simply to differentiate trending or non-trending price behavior.

Ichimoku Cloud Charts

Ichimoku charts were devised in the 1930s in Japan. While the charts are complex to look at for the uninitiated, they are great for trend identification once they are learned and understood. Ichimoku translates to 'one look' and the value of these cloud charts is the trend can be assessed at a glance. The basic idea is price above the cloud is bullish and below the cloud is bearish. The shaded cloud area effectively infills between two averages. The advantage here is that the width of cloud provides an area of price uncertainty rather than just a single line. This provides scope to consider that the trend is neutral when prices are in the cloud.

This is a brief summary of how to use Ichimoku clouds for trend identification. It is by no means exhaustive and there are many nuances that are beyond the scope of this paper.

There are five lines, each generated from price, on an Ichimoku chart as follows:

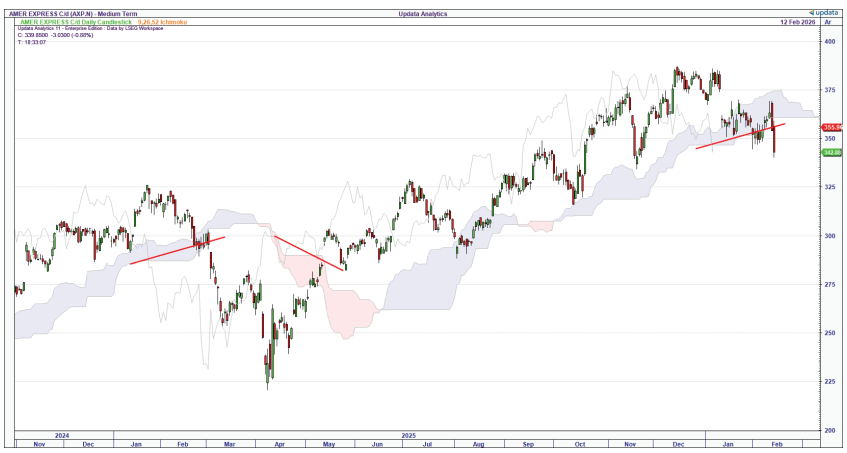
- The turning line – a midpoint average of the last 9 bars
- The standard Line – a midpoint average of the last 26 bars
- Cloud Span A – a midpoint of the turning and standard line values, shifted forward 26 bars
- Cloud Span B – a midpoint of the last 52 bars, shifted forward by 26 bars
- The lagging line – each bar's closing value shifted back by 26 bars

The turning and standard lines can be used in a similar way as moving averages crossing and prices will often find support or resistance on these lines as well. These averages are looking at the midpoint of trading across their period. This makes much more sense in 24-hour traded markets where the close assumed at midnight is purely arbitrary and not the close at all. What matters is the extreme points of the trading range. In sideways markets especially, prices in the cloud are trading in the range.

The main value of these first two lines is in their role in constructing the outer bands of the cloud. The cloud is effectively a combination of midpoint averages shifted ahead of the price. The Lagging line is the line of closing prices shifted back 26 bars. Many analysts find that the clouds are too noisy with the turning and standard lines drawn on the chart and find it best to remove them. The lagging line can also be toned down to a fainter line so that price and the cloud, the most important aspects of the Ichimoku chart, can be studied with more clarity. Figure 6 shows an Ichimoku chart of IBM with the first two midpoint averages removed and the lagging line shown as a fainter line making the chart easier to read.

The cloud forms a highly visual shaded area of support and resistance between the cloud spans on the chart. The simple interpretation is that if prices are above the cloud the trend is bullish and if they are below the cloud the trend is bearish. The cloud runs ahead of prices in the future and provides an area of support and resistance, rather than the more definite boundary of a line such as a moving average or bands.

Figure 6 - Daily Ichimoku chart of IBM with the first two midpoint averages removed. Chart: Updata Analytics, Data: LSEG



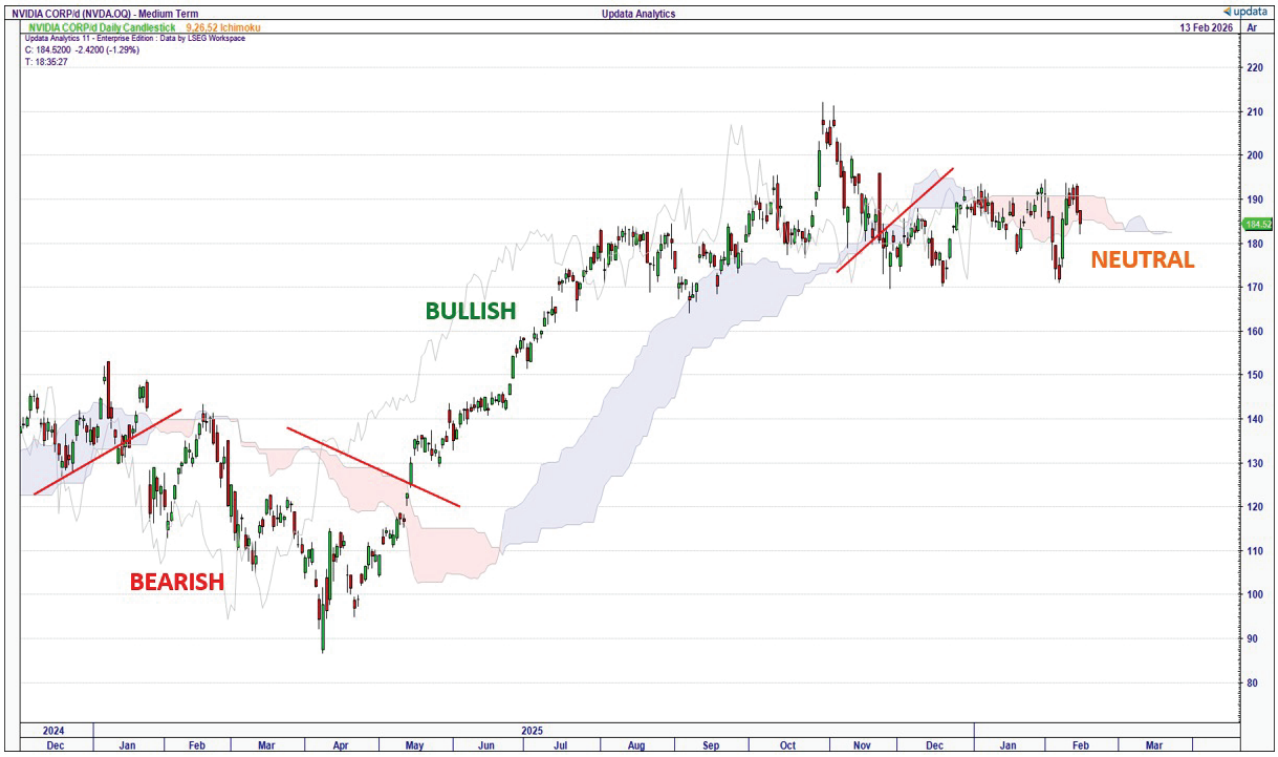
Prices will likely find support in a cloud when they enter form above. In this instance the cloud will likely have the faster moving Span A above the slower span B and the cloud will be pointing upwards. Most systems shade the cloud in a bullish color such as blue here. The cloud edges, the top and bottom of the cloud, will often provide more precise lines of support in uptrends. If prices entered the cloud from above, then support is being tested and the trend is still bullish.

In the case of a downtrend, prices will likely find resistance in a cloud when they enter form below. In this instance the cloud will likely have the faster moving Span A below the slower span B and the cloud will be pointing downwards. The cloud is normally shaded in a bearish color such as pink here. The cloud edges, the top and bottom of the cloud, will often provide more precise lines of resistance in downtrends. If prices entered the cloud from below, then resistance is being tested and the trend is still bearish.

If a full transition from one side of the cloud to the other occurs, this marks a trend change. A move in prices from below the cloud to above the cloud marks a change in trend from bearish to bullish. The cloud, that was previously falling, provided an area of resistance will now be rising with rising prices and now provide an area of price support. Conversely, a move in prices from above the cloud to below the cloud marks a change in trend from bullish to bearish. The cloud, that was previously rising, provided an area of support will now be falling with falling prices and now provide an area of price resistance.

While the settings can be changed for the cloud construction, it is not recommended. Backtests show there is no advantage in doing this. The big advantage of the technique is that it is a highly visual mechanism for establishing if a chart is bullish, bearish or neutral in an instant. Figure 7 shows a daily chart of Nvidia with the trends marked on the chart.

Figure 7 -Daily Ichimoku chart of Nvidia with trends marked



The Ichimoku cloud technique is trend following method of technical analysis which retrospectively signals tops and bottoms. Because it is trend following the charts are less clear in choppy rangebound markets. If the cloud is noisy in this way, then the trend is not up or down, but sideways and we have to wait for a break higher or lower from the sideways trading range for a new uptrend or downtrend to show itself against the cloud accordingly. It is frequently the case, that if prices are trading sideways this is the neutral phase between trends changing from bullish to bearish or vice versa. This signals a time of caution on that chart time horizon.

Multiple Time Frame Analysis

Because the trend can be assimilated visually very quickly, swapping the chart frames from daily to weekly and back to daily to hourly gives the trend states for medium term, long term and short-term time horizons respectively. Figure 8 shows Long Term Weekly, Medium Term Daily, Short Term Hourly charts for 3M.

Figure 8 — 3M (MM) — Long Term Weekly, Medium Term Daily, Short Term Hourly charts



The long-term trend is clearly bullish with prices above the cloud. The assumption here is that the view of at least several months ahead is bullish. The medium-term trend, with prices in the cloud, is neutral and the short-term trend is clearly bearish. Investment decisions from a technical analysis perspective for this example will depend on time horizon. The weekly cloud base is around $150, and rising, giving quite a margin of pullback in price to live with. A pullback in price of this magnitude occurred in 2025 all the way back to the cloud before recovering to make new highs. Looking at the medium-term daily chart, prices are in a sideways consolidation phase. There is a support level from previous price lows around $150 as well. A move below here and the medium-term trend would be more clearly bearish below the cloud. This support level is also marked on the weekly chart and here, if a new high is not reached, a bearish top pattern would be given. The long-term weekly trend would still be bullish above $150, but the medium-term trend would be entering a clearer bear trend implying further falls in price in the weeks ahead. Again, the decision to ride out such falls would be determined by time horizon. The short term 60-minute chart turned bearish when the price fell through the cloud and a move below the current support level of $163 would be more bearish and a signal to watch the support levels on the medium-term daily chart. This example highlights the fact that using just one chart is not normally enough to make trading decisions. What can be said about 3M from these charts is that the stock is long term bullish, medium-term neutral, short term bearish.

Another unique aspect of Ichimoku cloud charts is that, because of the construction, the cloud extends ahead of the price. The averages often cross ahead of the price crossing the cloud which also acts as a warning and the cloud provides a 'thinking time' into the future. This provides a roadmap of where prices need to be in the future to maintain trend. Figure 9 shows the time horizon table with cloud extension for each time frame.

Figure 9 shows - time horizon table with cloud extension for each time frame.

	ULTRA SHORT	VERY SHORT	SHORT TERM	MEDIUM TERM	LONG TERM	VERY LONG	ULTRA LONG
Time Horizon	Minutes	Hours	Days	Weeks	Months	Years	Many Years
Chart Frame	tick/1 Min	5/10 Mins	Hourly	Daily	Weekly	Monthly	Quarterly
Cloud Extends	30 Mins	2-4 Hours	3 Days	1 Month	6 Months	2 Years	8 Years

The long-term trend is primary

As Charles Dow pointed out, over century ago, the longer-term trend is the dominant trend. This means that medium-term and short-term trends will normally be skewed in line with the long-term trend. The longer-term prevailing trends will normally have a bigger impact on the shorter-term trends and the shorter-term trends in price are usually just moves within the longer-term trend. But all trends eventually come to an end and in the case of the longer term turning, the shorter-term trends will have turned first. Because the long-term trend is the primary dominant trend, it pays, when conducting chart analysis, to look at the longer-term chart first.

Figure 10 — start with the weekly long-term chart, then medium-term daily, then short-term



Figure 10 shows the process of starting with the long-term weekly chart, before moving to the medium-term daily and then on to the short-term 60-minute chart. Once the long term-trend is clearly established, it is possible to swap back and forth looking at the medium-term and short-term trends. This process is similar to using a zoom lens on a camera. In order to know what to zoom into, it is necessary to have a wider depth of field to begin with.

Scoring the trends

While it is not normally possible to say with conviction that an instrument is bullish or bearish, measuring each of the different time frame trends does allow for a scoring system. With just these three time frames and three trend states to consider there are twenty-seven possible combinations. Figure 11 shows these variations. This highlights the complexity of defining the overall trend position when multiple trends on different time horizons may be in play. A method for defining overall trend strength is feasible by assigning more weight to the primary long-term trend and less to the short term to give an overall score. Giving positive scores for bullish trends, zero for neutral and negative scores for bearish trends as well provides one potential mechanism for defining overall trend. This method could be applied to other trend definition methods across the time frames, but the visual and objective quality of Ichimoku charts makes it the quicker and easier to establish than other techniques.

Figure 11 — A mechanism for measuring overall trend strength

Long Term	Medium Term	Short Term	Score			Total
Bullish	Bullish	Bullish	3	2	1	6
Bullish	Bullish	Neutral	3	2	0	5
Bullish	Neutral	Bullish	3	0	1	4
Bullish	Bullish	Bearish	3	2	-1	4
Bullish	Neutral	Neutral	3	0	0	3
Neutral	Bullish	Bullish	0	2	1	3
Bullish	Neutral	Bearish	3	0	-1	2
Neutral	Bullish	Neutral	0	2	0	2
Bullish	Bearish	Bullish	3	-2	1	2
Neutral	Neutral	Bullish	0	0	1	1
Neutral	Bullish	Bearish	0	2	-1	1
Bullish	Bearish	Neutral	3	-2	0	1
Bullish	Bearish	Bearish	3	-2	-1	0
Neutral	Neutral	Neutral	0	0	0	0
Bearish	Bullish	Bullish	-3	2	1	0
Bearish	Bullish	Neutral	-3	2	0	-1
Neutral	Bearish	Bullish	0	-2	1	-1
Neutral	Neutral	Bearish	0	0	-1	-1
Bearish	Bullish	Bearish	-3	2	-1	-2
Neutral	Bearish	Neutral	0	-2	0	-2
Bearish	Neutral	Bullish	-3	0	1	-2
Neutral	Bearish	Bearish	0	-2	-1	-3
Bearish	Neutral	Neutral	-3	0	0	-3
Bearish	Bearish	Bullish	-3	-2	1	-4
Bearish	Neutral	Bearish	-3	0	-1	-4
Bearish	Bearish	Neutral	-3	-2	0	-5
Bearish	Bearish	Bearish	-3	-2	-1	-6

This table uses a traffic light system and change could be shown each time the table is updated, using a bold font to indicate change for instance. When the trends align, such as bullish, bullish, bullish or bearish, bearish, bearish a stronger conviction is achieved as to whether prices will rise or fall respectively. When the trends are not aligned prices may fall before they rise or rise before the fall making price prediction more difficult. The position of price in relation to the cloud is easily determined such that a table for any list of instruments can be easily be generated with some simple programming either in software or with Excel formulas. Figure 11 shows, such a table for the Dow Jones Index constituents.

Figure 11 - Table for the Dow Jones Index constituents. Source: Updata Analytics



Dow 30 - Trend Sates

Name	Code	Mid	+/- Chg	% Chg	Long Term(W)	Medium Term	Short Term(60)	Tot
JOHNSON JOHNSO/d	JNJ.N	246.61	4.12	1.70	Bullish	Bullish	Bullish	6
MCDONALD'S COR/d	MCD.N	332.53	3.30	1.00	Bullish	Bullish	Bullish	6
THE TRAVELERS /d	TRV.N	304.44	-0.49	-0.16	Bullish	Bullish	Bullish	6
VERIZON COMMS/d	VZ.N	50.29	1.04	2.11	Bullish	Bullish	Bullish	6
HNYWL INTRTL O/d	HON.OQ	242.90	-1.07	-0.44	Bullish	Bullish	Bullish	6
COCA-COLA CO/d	KO.N	80.47	0.63	0.79	Bullish	Bullish	Bullish	6
NVIDIA CORP/d	NVDA.OQ	190.68	0.85	0.45	Bullish	Bullish	Bullish	6
AMGEN/d	AMGN.OQ	378.42	3.67	0.98	Bullish	Bullish	Bullish	6
CHEVRON/d	CVX.N	184.81	0.88	0.48	Bullish	Bullish	Bullish	6
MERCK & CO/d	MRK.N	123.47	1.21	0.99	Bullish	Bullish	Bullish	6
CISCO SYSTEMS/d	CSCO.OQ	77.55	-1.65	-2.08	Bullish	Bullish	Neutral	5
CATERPILLAR IN/d	CAT.N	758.96	-0.78	-0.10	Bullish	Bullish	Neutral	5
WALMART INC OR/d	WMT.OQ	125.61	2.62	2.13	Bullish	Bullish	Bearish	4
BOEING CO/d	BA.N	229.35	-2.68	-1.16	Bullish	Bullish	Bearish	4
SHERWIN-WILLIA/d	SHW.N	361.31	0.68	0.19	Bullish	Bullish	Bearish	4
PROCTER & GAMB/d	PG.N	164.27	3.49	2.17	Neutral	Bullish	Bullish	3
3M COMPANY/d	MMM.N	165.75	-1.31	-0.78	Bullish	Neutral	Neutral	3
GOLDM SACHS GR/d	GS.N	893.41	-28.83	-3.13	Bullish	Neutral	Bearish	2
APPLE INC/d	AAPL.OQ	266.22	1.64	0.62	Bullish	Bearish	Neutral	1
JPMORGAN CHASE/d	JPM.N	298.79	-12.00	-3.86	Bullish	Bearish	Bearish	0
AMER EXPRESS C/d	AXP.N	321.22	-24.96	-7.21	Bullish	Bearish	Bearish	0
AMAZON COM/d	AMZN.OQ	204.11	-6.00	-2.86	Neutral	Bearish	Neutral	-2
HOME DEPOT INC/d	HD.N	374.72	-7.53	-1.97	Bearish	Bullish	Bearish	-2
INTL BUS MACHI/d	IBM.N	247.06	-10.10	-3.93	Neutral	Bearish	Bearish	-3
VISA INC/d	V.N	311.72	-9.23	-2.88	Neutral	Bearish	Bearish	-3
WALT DISNEY CO/d	DIS.N	103.53	-2.05	-1.94	Neutral	Bearish	Bearish	-3
UNITEDHEALTH G/d	UNH.N	285.12	-4.88	-1.68	Bearish	Bearish	Neutral	-5
NIKE INC CL B/d	NKE.N	62.25	-3.15	-4.82	Bearish	Bearish	Bearish	-6
SALESFRC INC O/d	CRM.N	174.84	-10.32	-5.57	Bearish	Bearish	Bearish	-6
MICROSOFT CP/d	MSFT.OQ	386.66	-10.57	-2.66	Bearish	Bearish	Bearish	-6

The rules for defining trend state are:

- Bullish is when price and the lagging line are both above the cloud
- Bearish is when price and lagging line are both below the cloud
- Neutral is when either price or lagging line are in the cloud

These rules can easily be programmed into a spreadsheet using the 5 Ichimoku lines outlined and applied across a database of historical price data.

The time frames are scored as follows:

- Long Term Weekly Chart – Bullish – 3, Neutral = 0, Bearish = -3
- Medium Term Daily Chart – Bullish – 2, Neutral = 0, Bearish = -2
- Short Term Hourly Chart – Bullish – 1, Neutral = 0, Bearish = -1

This favours the longer-term trends and leads to a maximum positive score of 6, bullish on all time frames and a minimum score of -6, bearish on all time frames.

This methodology can be easily be applied to large universes of instruments which can then be ranked by clicking on the header of the total score column. This can sort an entire market, such as the Russell 3000 stocks from most bullish overall to most bearish.

The three time horizons could be changed, such as monthly, weekly and daily for a very long outlook or to daily, 60-minute and 10-minute charts for a shorter-term outlook. Other technical analysis criteria for defining trend, such as those already mentioned, could be used for scoring. More criteria could be added and ranking could be more granular using percentages. For example, Relative strength against an index, normalized to 100 from a given date and weighted to the time frames would give a wider range of scoring across a universe of stocks.

The key aspect of this scoring mechanism is the measure runs across multiple time frames to arrive at an overall view of instruments with a portfolio and across the market.

Trading such a system

In terms of making trading and investment decisions, a fund could set rules based around overall scores and most importantly how the scores are changing. For example, a long only fund may set a rule that the long-term trend always has to be bullish for all investments held. Looking at Figure 11, the top 5 entries, with a score of 3 or greater, would qualify. The sixth entry; Neutral, Bullish, Bullish, which also has a s score of 3, could be considered an opportunity to enter an investment on the basis that the shorter-term trends could drive the longer-term trend bullish as well. There are four other entries further down the table with a bullish long- term trend, but the shorter-term trends contain a bearish state, so this would not be tolerated for any long positions a fund might hold. These rules could be inverted for short trading, where the score would need to be between -3 and -6.

Assigning the largest score to the longer-term trend is in line with the first and sixth rules of Dow Theory, where the longer-term trend is primary and will have influence over the secondary and tertiary trends.

Position sizing and legging in and out of trades

The sweet spot for entering a long position is probably a state of Bullish, Bullish, Neutral where the short-term trend has turned from bearish. A trade could be entered on the move to short-term neutral from short-term bearish on the basis that the longer-term bullish trends will likely impact the short-term trend. More certainty is likely when the shorter-term trend changes to bullish, but this will likely be less profitable as prices will have moved higher. Each scenario could be position sized differently presenting the possibility to leg into a trade.

Holdings could also be reduced on the basis of scores and how the trends are changing. The top two scores of 6 and 5 are likely no cause for action for a long only fund. The move to 4 could present a case for weighting a holding down, to 3 reducing a holding further. A move below 3 would be cause for exiting a holding completely. Again, for short trading, the inverse scores -6 to -3 would be for short positions. Note that the scores from +2 to -2 are the instruments where the trend is not strong and there is a higher likelihood of underperformance.

The legging in and out of trades in the manner described above is in line with the second rule Dow Theory in an attempt to capture the accumulation phase of a trend on entering long and distribution phase for long exit. The objective is to be in ahead of the participation phase, where the wider market becomes involved, and to exit before the wider market exits.

Money Management

Capital preservation is king for any fund and this scoring strategy ensures that the strongest trending instruments are held. There will be a minority of instances where it would have paid to wait it out for an instrument with a deteriorating score below 3 but there will also be many other opportunities to enter instruments with rising scores above 3. Here the likelihood of out-performance on average will be greater.

This strategy ensures that the strongest trending instruments across multiple time horizons are held.

CFTe
IFTA Certified Financial Technician

Certified Financial Technician (CFTe) Program

IFTA Certified Financial Technician (CFTe) consists of the CFTe I and CFTe II examinations. Successful completion of both examinations culminates in the award of the CFTe, an internationally recognised professional qualification in technical analysis.

Examinations

The **CFTe I** exam is multiple-choice, covering a wide range of technical knowledge and understanding of the principals of technical analysis; it is offered in English, French, German, Italian, Spanish, Arabic, and Chinese; it's available, year-round, at testing centers throughout the world, from IFTA's computer-based testing provider, Pearson VUE.

The **CFTe II** exam incorporates a number of questions that require essaybased, analysis responses. The candidate needs to demonstrate a depth of knowledge and experience in applying various methods of technical analysis. The candidate is provided with current charts covering one specific market (often an equity) to be analysed, as though for a Fund Manager.

The CFTe II is also offered in English, French, German, Italian, Spanish, Arabic, and Chinese, via Zoom, typically in April and October of each year.

Curriculum

The CFTe II program is designed for self-study, however, IFTA will also be happy to assist in finding qualified trainers. Local societies may offer preparatory courses to assist potential candidates. Syllabuses, Study Guides and registration are all available on the **IFTA website.**

To Register

Please visit our website for **registration details.**

Cost

IFTA Member Colleagues	Non-Members
CFTe I $550 US	**CFTe I $850 US**
CFTe II $850* US	**CFTe II $1,150* US**

**Additional Fee (CFTe II only): $100 US Proctor Fee*

MFTA
IFTA Master of Financial Technical Analysis

Master of Financial Technical Analysis (MFTA) Program

IFTA's Master of Financial Technical Analysis (MFTA) represents the highest professional achievement in the technical analysis community, worldwide. Achieving this level of certification requires you to submit an original body of research in the discipline of international technical analysis, which should be of practical application.

Examinations

In order to complete the MFTA and receive your Diploma, you must write a research paper of no less than three thousand, and no more than five thousand, words. Charts, Figures and Tables may be presented in addition.

Your paper must meet the following criteria:

- It must be original
- It must develop a reasoned and logical argument and lead to a sound conclusion, supported by the tests, studies and analysis contained in the paper
- The subject matter should be of practical application
- It should add to the body of knowledge in the discipline of international technical analysis

Timelines & Schedules

There are two MFTA sessions per year, with the following deadlines:

SESSION 1	
"Alternative Path" application deadline	**February 28**
Application, outline and fees deadline	**May 2**
Paper submission deadline	**October 15**

SESSION 2	
"Alternative Path" application deadline	**July 31**
Application, outline and fees deadline	**October 2**
Paper submission deadline	**March 15 (of the following year)**

To Register

Please **visit our website** for further details and to register.

Cost

$950 US (IFTA Member Colleagues);
$1,200 US (Non-Members)

IFTA
International Federation of Technical Analysts

Prediction is Very Difficult: Especially if it's About the Financial Markets! (Part III)

A rigorous evaluation of statistical forecasting methods and their comparative effectiveness.

Davide Pandini, PhD, MFTA[1], CFTe[2], CMT[3], CSTA[4]

Ronco Briantino, Monza e Brianza, Italy
davide.pandini@hotmail.com

"... Prediction is very difficult, especially if it's about the future! ..."
Niels Bohr (1885-1962), Danish physicist, Nobel Prize in Physics (1922)

Abstract

Accurately predicting asset prices and market trends in the dynamic landscape of financial markets is an open challenge. This article is the third part of a series dedicated to a thorough exploration, review, and rigorous evaluation of statistical forecasting methods for financial time series. The first part [1], published in the IFTA Journal 2025, introduced the evaluation framework of the series: the fundamental structure of the forecasting models, the historical dataset and its segmentation into training and testing ranges, the forecasting accuracy metrics and the information loss criteria, and the simple forecasting methods that serve as baselines for the more sophisticated techniques. The second part [2], published in the IFTA Journal 2026, presented a comprehensive analysis of the exponential smoothing methods, from the Brown simple exponential smoothing to the Holt, Gardner, Taylor, and Holt-Winters families of models. This third part is dedicated to the AutoRegressive Integrated Moving Average (ARIMA) models and to their seasonal extension (SARIMA), which provide an explicit and interpretable representation of the conditional mean of a financial time series.

The study adopts a rigorous analytical framework, incorporating hypothesis testing, verification, and thorough model selection and validation processes. The assessment of first-order and seasonal stationarity, through the deterministic trend, Phillips-Perron, and deterministic seasonality tests, guides the specification of the differencing orders, while the autocorrelation and partial autocorrelation functions support the identification of the autoregressive and moving average orders. The most significant ARIMA and SARIMA models are then derived from a general template, optimally estimated within the training range, and selected according to the Akaike and Bayesian information criteria. Their forecasting accuracy is systematically evaluated in the independent testing range, considering both multistep forecasting and one-step forecasting without re-estimation, while the Ljung-Box test is employed to assess the white noise property of the models' residuals, and hence the adequacy of the selected models.

The results presented in this work confirm the value of the ARIMA models as an interpretable and statistically disciplined toolkit for short-term forecasting of financial asset prices. At the same time, the residuals diagnostics reveal the limitations intrinsic to a modeling framework that assumes a constant variance, and thus motivate the fourth and final part of the series, already planned for the IFTA Journal 2028, which will be dedicated to the advanced forecasting models: the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models, the hybrid ARIMA-GARCH framework, the threshold GJR-GARCH extension, and the non-Gaussian GARCH models based on the Student's t distribution.

1. Introduction and Motivation

The financial markets and the prices of the asset classes are complex systems influenced by several factors, ranging from economic and financial indicators to global geopolitical events. Therefore, the ability to accurately predict asset prices and market trends remains a paramount challenge of critical interest for stakeholders such as technical and fundamental analysts, traders, and investors. To address this challenge, this work provides a comprehensive exploration and a rigorous assessment of statistical forecasting methods and of their comparative effectiveness in capturing the complexity of financial market dynamics.

This article is the third part of a series dedicated to the rigorous evaluation of statistical forecasting methods for financial time series. The first part [1], published in the IFTA Journal 2025, introduced the motivation and the general framework of the study: the fundamental structure of forecasting models (additive and multiplicative), the historical dataset and its segmentation into training and testing ranges, the forecasting accuracy metrics and model selection criteria (RMSE, MAPE, MASE, AIC, and BIC), and the simple forecasting methods (arithmetic mean, random walk, seasonal random walk, and random walk with drift) that serve as baselines for the more sophisticated techniques. The second part [2], published in the IFTA Journal 2026, presented a comprehensive analysis of the exponential smoothing methods, from the Brown simple exponential smoothing to the Holt, Gardner, Taylor, and Holt-Winters families of models, together with the procedures for optimal parameter estimation and model selection.

In this issue of the IFTA Journal the series continues with the AutoRegressive Integrated Moving Average (ARIMA) models and their seasonal extension (SARIMA), which constitute the reference framework for modeling the conditional mean of a financial time series. The series will be completed in the IFTA Journal 2028 with the fourth and final part, dedicated to the advanced forecasting models: the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models, the hybrid ARIMA-GARCH framework, the threshold GJR-GARCH extension, and the non-Gaussian GARCH models based on the Student's t distribution.

The motivation underlying the entire series remains central to this part. Data-driven approaches, such as machine learning and neural networks, may have the capability of handling large datasets and uncovering involved patterns, but they often lack the interpretability and rigor of statistical models, and they are prone to overfitting. Statistical methods, conversely, provide a structured and proven framework for financial forecasting: their parameters and assumptions are explicitly defined, their outputs can be clearly understood and reasoned about by analysts, traders, and investors, and their capacity to adapt to the underlying assumptions of financial data allows them to effectively capture the temporal dynamics inherent in financial assets. These properties are particularly evident in the models discussed in this part of the series. ARIMA models are grounded in the concepts of stationarity, differencing, and autocorrelation analysis, and each of their components has a transparent interpretation: the autoregressive terms describe the dependency on the past values of the time series, the differencing order removes the trend and the seasonality, and the moving average terms account for the past forecasting errors. Moreover, as we will show, the ARIMA framework is general enough to encompass, as particular cases, several of the simple and exponential smoothing methods analyzed in the previous parts of the series, thus providing a unifying perspective on the models presented so far.

For consistency with the previous parts of the series, and to allow a direct comparison of the results, the same historical dataset is used throughout this work. The data consists of the S&P 500 State Street Exchange Traded Fund daily adjusted close prices (ticker: SPY), spanning about nineteen business calendar years, from December 1st, 2004, to July 31st, 2023, for a total of 4,697 observations. The adjusted close prices account for corporate actions such as dividends and stock splits, ensuring that the historical price data are consistent and comparable over time. The time series is divided into a training range for optimal parameter estimation and fine tuning, covering 3,424 observations from December 1st, 2004, to July 9th, 2018, and a testing range for forecasting accuracy and robustness evaluation, including 1,273 observations from July 10th, 2018, to July 31st, 2023. This historical window was selected because it encompasses multiple bull and bear markets, including the global financial crisis (2007–2008), the COVID-19 sell-off (March 2020), and the bull market lasting over a decade, thus enabling the assessment of the forecasting models across significantly different market conditions.

As in the previous parts of the series, the evaluation follows a rigorous statistical methodology grounded in hypothesis testing, verification, and meticulous model selection and validation within the training range, followed by systematic testing in the independent testing range, considering both multistep forecasting and one-step forecasting without re-estimation. In this part, the methodology is extended with the statistical tools required by the ARIMA framework: the assessment of first-order and seasonal stationarity through the deterministic trend, Phillips-Perron, and deterministic seasonality tests; the specification of the model orders through the autocorrelation and partial autocorrelation functions; the estimation of the optimal model parameters; the selection of the models with the least information loss through the Akaike and Bayesian information criteria; and the verification of the white noise property of the residuals through the Ljung-Box test, which assesses whether the selected model has adequately captured the information contained in the historical data.

1.1 Structure of the Paper

This article is organized as follows: Section 2 introduces both non-seasonal and seasonal ARIMA models, delving into the critical concept of time series stationarity, describing the estimation of the optimal model parameters, and deriving the specific ARIMA and SARIMA models from a general template. In Section 3, the most significant ARIMA models are optimized within the training range, selected through the information loss criteria, and their forecasting accuracy is evaluated in the testing range for both multistep forecasting and one-step forecasting without re-estimation; the Ljung-Box test is then employed to assess the white noise property of the models' residuals and, consequently, the adequacy of the selected models. Finally, Section 4 outlines the concluding remarks of this article, summarizing the key findings and introducing the topics of the fourth and final part of the series.

A central contribution of this third part of the series is the thorough analysis and evaluation of the ARIMA and SARIMA models, presented not as a catalogue of formulas, but as a structured methodological path in which every modeling decision — the differencing order, the autoregressive and moving average orders, the inclusion of a constant drift, the seasonal components — is motivated by the verification of an explicit statistical hypothesis. The paper meticulously guides readers through the essential steps required for a statistically rigorous assessment of the presented models and for their correct application to the forecasting of financial time series. While mathematical formulas are included for a comprehensive understanding of how to control and interpret the parameters and verify the assumptions, readers can navigate the paper by focusing only on the statistical methodology employed to assess the models and on the experimental results, bypassing the mathematical details. This article is self-contained, although readers are referred to the previous two issues of the series [1][2] for the complete treatment of the evaluation framework, of the simple forecasting methods, and of the exponential smoothing models.

2. Auto Regressive Integrated Moving Average Models

Autoregressive integrated moving average (ARIMA) models are commonly used time series forecasting techniques that combine three components: autoregressive (AR), integrated (I), and moving average (MA), and are denoted as ARIMA(p,d,q). These components can be summarized as follows:

- *Autoregressive (AR) component:* This component captures the relationship between the current value of the time series and its past values. It assumes that the current value can be predicted by linearly combining the previous values with

some multiplicative coefficients. The (p) parameter in ARIMA represents the number of lagged terms (past observations) used in the autoregressive component;
- *Integrated (I) component:* The integration component refers to differencing the time series to make it stationary. Stationarity is important because many time series models assume that the statistical properties of the data do not change over time. Differencing involves taking the difference between consecutive observations to remove any trend or seasonality. The (d) parameter in ARIMA represents the number of differencing operations needed to achieve stationarity;
- *Moving average (MA) component:* The moving average component represents the relationship between the current value of the time series and its past forecast errors (i.e., residuals). It accounts for short-term fluctuations that are not explained by the autoregressive and differencing components. The (q) parameter in ARIMA represents the number of lagged forecast errors or residuals used in the moving average component.

An ARIMA(p,d,q) model forecasts a time series by considering its past values (autoregressive), making it stationary if necessary (integration or differencing), and accounting for past forecast errors (moving average). For example, an ARIMA(1,1,1) model implies:

- AR(1): The current value is predicted based on the previous value, i.e., one lagged term;
- I(1): The time series is differenced once to achieve stationarity;
- MA(1): The model uses one lagged forecast error or residual to capture short-term fluctuations.

The order of the ARIMA components (p,d,q) can vary depending on the characteristics of the time series. Selecting appropriate values for these parameters often involves analyzing the autocorrelation and partial autocorrelation plots of the data. Although ARIMA models are versatile and widely used in forecasting various types of time series data, nevertheless, ARIMA models assume linearity, stationarity, and that the relationships among the variables are fixed over time. When these assumptions are not met, more advanced models such as SARIMA (Seasonal ARIMA) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models may be considered. In summary, ARIMA models consist of modeling the conditional mean of a stochastic process, where the "conditional mean" refers to the expected value of a future observation given its past values and possibly past forecast errors. The model estimates the (p,d,q) parameters that best fit the observed data, to make forecasts at future time points. The golden reference for ARIMA is [6]. This classic text, first published in 1970 and later updated in subsequent editions [3], is considered one of the fundamental works in time series analysis and covers the theory and application of ARIMA models in detail.

A *seasonal autoregressive integrated moving average* (SARIMA) model is a time series forecasting model that extends the ARIMA model to account for both non-seasonal and seasonal patterns in time series data. SARIMA models are used to make forecasts and capture the complex dependencies within data that exhibit seasonal fluctuations. The SARIMA(P,D,Q)m model accounts for seasonal patterns by introducing a seasonal autoregressive parameter (P) that models the relationship between the current observation and observations from the same season in previous periods. The (P) parameter represents the order of the seasonal autoregression, indicating how many seasons are considered. The (D) parameter represents the seasonal differencing required to make the seasonal part of the data stationary. The (Q) parameter models the relationship between the current observation and past seasonal forecast errors and represents the order of the seasonal moving average. Furthermore, SARIMA models have a seasonal period (m) that indicates the number of time periods in one season. For example, when modeling daily data with monthly seasonality, (m) would be 21. In summary, a SARIMA model is defined by the following parameters: (p, d, q) for the non-seasonal part and (P, D, Q, m) for the seasonal part. These parameters are determined through data analysis and the examination of Seasonal Autocorrelation Function (SACF) and Seasonal Partial Autocorrelation Function (SPACF) plots.

2.1 Optimal Parameters Estimation

The optimal parameter estimation in ARIMA models consists of finding the values of the autoregressive order (p), differencing order (d), and moving average order (q) resulting in the best-fitting model for a given time series. The optimal ARIMA parameters can be estimated by means of:

- *Visual inspection:* One of the simplest methods is to visually inspect the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the time series. The ACF and PACF plots can help identify potential values for (p) and (q) based on the significant spikes in the plots. A lag where the ACF drops off significantly after initially high values might suggest the order of the moving average component (q), while a lag at which the PACF drops off significantly could indicate the order of the autoregressive component (p);
- *Automated methods:* There are automated algorithms and functions available in software packages that can help identify optimal parameters. In this work, we used the Excel solver package to find the best fitting ARIMA models based on the minimization of the objective function SSE (sum of squared forecast errors or residuals);
- *Cross-validation:* Cross-validation techniques involve dividing the data into training and testing datasets. We fit ARIMA models with different parameter combinations on the training data and assess their performance on the testing data using metrics like Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE);
- *Information criteria:* Criteria like AIC [4] and BIC [5] are quantitative measures that balance model fit and complexity. These criteria consider both the goodness of fit and the number of parameters in the model. Lower values of AIC or BIC indicate a better model fit.

The process of parameter estimation in ARIMA models involves a trade-off between model complexity and goodness of fit. An overly complex model may overfit the data and perform poorly on new observations, while an overly simplistic model may fail to capture important patterns in the data. The goal is to select parameters resulting in a model that generalizes well to unseen data, while adequately capturing the underlying dynamics of the time series.

Optimal parameter estimation first estimates and then fine tunes coefficients which minimize the sum of squared residuals (i.e., SSE) or forecast errors within the training dataset. Optimal parameter estimation in ARIMA models is achieved by means of two options. The first option is linear regression in which the d-order differenced current period data is given by the expression:

$$Linear\ Regression\ ARIMA(p,d,0): \Delta^d y_t = \alpha + \sum_{i=1}^{p} \beta_i * \Delta^d y_{t-i} + e_t$$

where (p) is the autoregressive order, (d) is the order of integration or differencing, (0) is the moving average order. The second option is based on Excel solver to minimize the objective function SSE:

$$Solver\ Objective\ ARIMA(p,d,q): \min\left(\sum_{t=1}^{n} e_t^2\right)$$
$$Solver\ Coefficients\ ARIMA(p,d,q): \alpha, \beta_p, \gamma_q$$
$$e_t = y_t - \hat{y}_t$$

where the forecast error or residual at current period e_t is given by the difference between the current period actual data y_t and the current period forecasted data $\hat{y}_t$. The estimated coefficients will be: the constant mean or drift α, the coefficients β_p, which correspond to the autoregressive component, and the coefficients γ_q, which correspond to the moving average component.

The standard error of the optimal parameters consists of the standard deviation estimation of the model residuals for each estimated or fine-tuned coefficient. For the ARIMA(p,d,q) model, it is calculated through linear regression and is given by:

$$se_{\alpha,\beta_p,\gamma_q} = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{(n-2) * \sum_{t=1}^{n} (x_t - \bar{x})^2}}; x = \Delta^d y_{t-p}, e_{t-q}$$

where $\bar{x}$ is the corresponding arithmetic mean and x corresponds to either the d-order differenced previous period data, or the corresponding previous period forecast errors or residuals, (p) and (q) are the autoregressive and moving average order respectively, and n is the number of observations.

Optimal parameters p-value estimation consists of determining whether the estimated or fine-tuned coefficients are individually statistically significant. We calculate the p-value through linear regression for the constant mean or drift α and for the autoregressive coefficient β_p and the moving average coefficient γ_q:

$$p_{\alpha,\beta_p,\gamma_q} = t_{dist,2\ tails}\left(abs\left(\frac{\alpha,\beta_p,\gamma_q}{se_{\alpha,\beta_p,\gamma_q}}\right), n-(p+q+P+Q+c)\right).$$

The p-value is given by the t distribution of a two-tailed test of the absolute values of the coefficients divided by their corresponding standard errors, and the number of degrees of freedom expressed as n (the number of observations) minus the number of all parameters, in which (p) is the autoregressive order, (q) is the moving average order, (P) is the seasonal autoregressive order, (Q) is the sea-sonal moving average order and (c) is a binary term corresponding to the constant mean or drift if the model includes it.

2.2 ARIMA Stationarity

A time series is first-order stationary, or weak stationary, when its statistical properties remain relatively constant over time. In other words, the mean and variance of the time series do not change as time progresses. This concept is fundamental in time series analysis, as many statistical modeling techniques assume stationarity. For a time series to be considered first-order stationary, the following conditions must be fulfilled:

- *Constant mean:* The mean of the time series remains the same for all time points;
- *Constant variance:* The variance of the time series remains constant for all time points;
- *Autocovariance independence:* The autocovariance between the time series data points y_t and y_{t+h} depends only on the time lag h and not on the specific time t. Mathematically, this property can be expressed as: $cov(y_t, y_{t+h}) \cong h$;
- *No systematic patterns:* There are no underlying systematic trends or patterns in the data, such as linear or quadratic trends or seasonality.

First-order stationarity is a relatively mild form of stationarity and is often considered as a prerequisite for applying the ARIMA models. It is important to note that not all time series are first-order stationary, and transformations or differencing may be necessary to achieve stationarity for analysis and modeling purposes.

Time series differencing is a technique used to transform the original sequence of data values into a new sequence that represents the changes between consecutive observations. In other words, it involves computing the difference or change between one data point and the next in a time-ordered sequence. The primary purpose of time series differencing is to stabilize the data and remove trends and seasonality. By differencing two consecutive values, the resulting series may become more stationary, which can make it easier to analyze the data by using various statistical methods and models. The process of time series differencing can be done at different orders denoted by the (d) parameter. For instance:

First-order differencing (d=1): This involves calculating the difference between each data point and the previous data point. It eliminates linear trends and can be used to stabilize variance;

Second-order differencing (d=2): This involves taking the difference between the first-order differences. It can be used to remove quadratic trends or seasonality.

The number of times we difference the time series is called

the order of differencing and depends on the properties of the time series. If the time series has a trend, we usually need a first-order differencing. If the time series has seasonality (or other higher-order patterns), we need to use second- or higher-order differencing. Although differencing can be a useful technique for improving the accuracy of time series forecasts, it is important to note that differencing can also introduce noise into the data. It is essential to be cautious when applying second-order differencing because it may lead to over-differencing, which by introducing unnecessary noise can result in an overly volatile time series. Over-differencing can make it more challenging to model and forecast the data accurately. This is because differencing amplifies the random fluctuations: over-differencing can lead to loss of relevant information, while under-differencing may not achieve the desired level of stationarity. The first-order differenced time series that can help achieve stationarity is expressed by the formula:

$$\Delta y_t = y_t - y_{t-1}$$

where Δy_t represents the current period first-order difference, y_t is the original time series, and y_{t-1} is the previous period time series. By subtracting the previous value from the current value, we obtain a new time series that represents the instantaneous rate of change or the increment between each pair of consecutive observations. This can help remove a linear trend, which is a step towards achieving stationarity. Second-order differencing is used to remove quadratic trends or patterns like seasonality. The formula for second-order differencing is given by:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2 * y_{t-1} + y_{t-2}$$

where $\Delta^2 y_t$ represents the current period second-order difference, while Δy_t and Δy_{t-1} represent the current and previous period first-order difference respectively. This formula captures the change in the rate of change of the time series. To sum up:

- *Removing linear trend:* On a time series with a linear trend (a gradual increase or decrease over time), we can use first-order differencing to remove the trend and make the data stationary;
- *Removing seasonality:* On a time series that exhibits seasonality, which is a repeating pattern at fixed intervals, we can use second-order differencing to remove the seasonality effect. The second-order differencing helps eliminate the periodic pattern by subtracting the differences between consecutive observations. An additional step for choosing the order of differencing to remove seasonality from the data is based on visual inspection of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series. The ACF and PACF can help identify the frequency of the seasonality. If the ACF and PACF show significant spikes at regular intervals, then the time series has seasonality at that frequency.

In a first-order trend stationary time series, the first-order differenced data is relatively stable and does not show any systematic increasing or decreasing patterns. This stationary data is often more responsive to ARIMA models. It is important to note that while first-order differencing can remove a linear trend, other underlying patterns like seasonality may still be present. Further analyses, such as examining ACF and PACF plots, might be necessary to determine the complete nature of the time series and select an appropriate modeling approach. To assess the first-order trend stationarity we can perform two tests:

1. Mean deterministic trend;
2. Phillips-Perron;

for both the level and the differenced time series. The mean deterministic trend test consists of a de-trending procedure based on subtracting the trend from the time series. The formula for current period data is:

$$y_t = \alpha + \beta * t + e_t$$

where the test for the β coefficient p-value consists of:

- If β coefficient p-value > 0.05 the time series is first-order trend stationarity with 95% statistical confidence;
- If β coefficient p-value < 0.05 then higher differencing order is needed for first-order trend stationary time series with 95% statistical confidence.

The Phillips-Perron test evaluates whether a time series is first-order trend stationary after first-order differencing, with the null hypothesis that it has a unit root and is not stationary (i.e., it contains a linear trend) [7]. It is a good choice for testing the presence of unit roots in time series regression models, which is an indication of non-stationarity. The alternative hypothesis is that the time series is stationary after first-order differencing (i.e., it has no unit root). The test consists in calculating the γ coefficient p-value from the following regression model:

$$\Delta y_t = \gamma * y_{t-1} + e_t$$

- If γ coefficient p-value < 0.05 the time series is first-order trend stationary with 95% statistical confidence;
- If γ coefficient p-value > 0.05 then higher differencing order is needed to achieve first-order trend stationarity with 95% statistical confidence.

In this work we evaluate first-order trend stationarity for the (adjusted) price time series (i.e., level time series) and the differenced time series through the deterministic trend test and the Phillips-Perron test. The numerical results for the level time series are shown in Table 1.

Table 1. Level time series

Deterministic Trend			Phillips-Perron		
Coefficient	Std Error	P-Value	Coefficient	Std Error	P-Value
0.04479	0.00041	0.00000	0.00040	0.00016	0.01145

Since the deterministic trend test p-value < 0.05, we can conclude that the price time series is not first-order trend stationary with 95% statistical confidence. To confirm this result, we performed the Phillips-Perron test. In this case the p-value is < 0.05 and we can reject the null hypothesis. Hence, according to Phillips-Perron the price time series is first-order trend stationary with 95% statistical confidence. We evaluate the deterministic trend test and the Phillips-Perron test on the first-order differenced time series. The numerical results are reported in Table 2.

Table 2. First-order differenced time series

Deterministic Trend			Phillips-Perron		
Coefficient	Std Error	P-Value	Coefficient	Std Error	P-Value
0.0000399	0.0000221	0.0707208	-1.0529593	0.0170761	0.0000000

The differenced time series has a deterministic trend test p-value > 0.05, therefore we can conclude that after first-order differencing the time series becomes first-order trend stationary. We perform the Phillips-Perron test on the first-order differenced time series and the p-value is < 0.05. We can conclude also with the Phillips Perron test that by performing first-order differencing the time series has become first order trend stationary with 95% statistical confidence.

2.3 SARIMA Stationarity

In a seasonal time series, the observations or measurements show a regular and predictable pattern or fluctuation at fixed intervals or periods. This seasonal pattern repeats over time and is often associated with specific calendar time units, such as days, weeks, months, quarters, or years. A first-order seasonal time series generally refers to a time series that exhibits a seasonal pattern or seasonality. The term first-order does not imply stationarity but rather refers to the presence of seasonality, typically without accounting for other non-stationary components like trends. In a first-order seasonal stationary time series, both regular (non-seasonal) differencing and seasonal differencing are applied to achieve stationarity. The first-order term implies that differencing has been applied once to make the data stationary, and this typically involves both regular differencing (d) and seasonal differencing (D):

- Regular Differencing (d): Regular differencing is applied to remove the non-seasonal trend-related component of the time series. Regular differencing subtracts from each observation the previous one. The number of regular differences needed to achieve stationarity is denoted as (d);
- Seasonal Differencing (D): Seasonal differencing is used to eliminate the seasonal component of the time series. The de-seasonalizing procedure involves subtracting previous season data from current period data depending on the seasonal pattern (year, month, quarter, etc.). The seasonally differenced data is equal to:

$$\Delta_m y_t = y_t - y_{t-m}$$
$$\Delta_m^2 y_t = \Delta_m y_t - \Delta_m y_{t-m}$$

Therefore, in a first-order seasonal stationary time series, it is common to have (d=1 first-order regular differencing) and (D=1 first-order seasonal differencing) to achieve both first-order stationarity and remove the seasonality. In a stationary series, the mean, variance, and autocorrelation structure do not change over time. Consequently, the seasonal pattern is no longer evident in the differenced series.

A mean deterministic seasonality test is a statistical test that checks whether the seasonal component of a time series is significant or not. It consists of a de-seasonalizing procedure based on subtracting seasonality from the time series by using dummy variables. Dummy variables are binary indicators that help capture the effects of categorical or qualitative factors, such as days, months, quarters, or seasons. The procedure for de-seasonalizing a time series using dummy variables can be summarized by the following steps:

1. *Data Collection and Identification of Seasonality*: Begin with the time series data and identify any seasonal patterns or fluctuations. Determine the periodicity of the seasonality (e.g., monthly, quarterly, yearly). In our work we consider a seasonality of 21 time periods;
2. *Create Dummy Variables*: Create dummy variables to represent the different seasons or time periods within the seasonal cycle. The number of dummy variables depends on the periodicity of the seasonality. For example: since we have seasonality on the monthly business cycle, we use 21 periods, each of those becoming a dummy variable;
3. *Fit a Regression Model:* Build a regression model where the dependent variable is the original time series, and the independent variables are the dummy variables created in step 2. The regression equation looks like this:

$$y_t = \sum_{i=1}^{m} \beta_i d_i + e_t$$

where m is the number of seasonal periods, β_i are the linear regression coefficients, d_i are the associated dummy variables, and et are the forecast errors or residuals;

4. *Estimate Model Coefficients:* Use regression techniques to estimate the coefficients (β_1, β_2, ... , β_m). These coefficients represent the average seasonal effect associated with each season or time period.

We test the β_i coefficients joint F-statistic p-value. This test is used to assess the collective significance of the seasonal effects represented by the dummy variables, and based on the p-value of the joint F-statistic we can conclude that:

- *Null Hypothesis:* The null hypothesis for the joint F-statistic test states that none of the coefficients for the dummy variables are significantly different from zero. In the context of de-seasonalization, this means that there is no significant seasonal effect in the time series;
- *Alternative Hypothesis:* The alternative hypothesis suggests that at least one of the coefficients for the dummy variables is significantly different from zero. This implies that there is a significant seasonal effect in the time series.

The conclusions based on the p-value associated with the joint F-statistic are:

- If the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. In this case, there is not enough evidence to conclude that any of the seasonal effects are significant, suggesting that the time series may not exhibit seasonality;
- If the p-value is less than the significance level of 0.05, we can reject the null hypothesis. This means that there is sufficient evidence to conclude that at least one of the seasonal effects is significant, and the time series exhibits seasonality.

In summary, by examining the p-value of the joint F-statistic, it is possible to determine whether the seasonal effects represented by the dummy variables are collectively significant or not. We performed the deterministic seasonality test for both level and seasonally differenced time series. The results are reported in Table 3 and Table 4 respectively.

Table 3. Level Time Series

Deterministic Seasonality	
F-Stat	P-Value
1092.6931	0.0000

Table 4. Seasonally Differenced Time Series

Deterministic Seasonality	
F-Stat	P-Value
7.3187	0.0000

The joint F-statistic is computed by means of linear regression on the dummy variables. The p-value is calculated by means of the F distribution where the degrees of freedom are given by the seasonality (i.e., 21), which in this case is also the regression degrees of freedom, while the residuals degrees of freedom is given by the total number of observations, which is the number of dependent or explained variables, minus the regression degrees of freedom, which is also the number of model parameters. The p-value in Table 3 is < 0.05, therefore the level time series is not seasonally stationary with 95% statistical confidence thus introducing the need of a higher order of seasonal differencing. We perform the deterministic seasonality trend test on the seasonally differenced data. Even in this case the p-value in Table 4 is < 0.05, therefore the time series is not first-order seasonally stationary with 95% statistical confidence.

2.4 ARIMA Model Specification

The ARIMA model specification consists of determining the parameters of the three main components: autoregressive (AR), integrated (I), and moving average (MA). The general expression for the ARIMA(p,d,q) model is given by the formula:

$$\Delta^d y_t = \alpha + \sum_{i=1}^{p} \beta_i * \Delta^d y_{t-i} + \sum_{j=1}^{q} \gamma_j * e_{t-j} + e_t.$$

For example: ARIMA(1,1,1) is expressed as:

$$\hat{y}_t = y_{t-1} + \beta_1 * (y_{t-1} - y_{t-2}) + \gamma_1 * e_{t-1},$$

while ARIMA(2,1,2)[C] has a constant term or drift —and is expressed as:

$$\hat{y}_t = \alpha + y_{t-1} + \beta_1 * (y_{t-1} - y_{t-2}) + \beta_2 * (y_{t-2} - y_{t-3}) + \gamma_1 * e_{t-1} + \gamma_2 * e_{t-2}.$$

Choosing the appropriate values for the parameters requires analysis of the data, including identifying trends, seasonality, and the order of differencing needed to achieve stationarity. In the ARIMA model specification we have differenced data and lagged differenced data to the number of periods in the forecasting seasonal lag. Since in this work we consider daily data on a business calendar, we approximate the monthly seasonality as 21 days, hence, we use a seasonal lag h=21 to compute the autocorrelation (ACF) and the partial autocorrelation (PACF) functions along with their corresponding upper and lower confidence intervals for 95% statistical confidence.

The autocorrelation function (ACF) measures the correlation between a time series and itself at different lags (i.e., time shifts), where the lag is the number of time periods between the two observations that are being correlated. In other words, it measures the similarity between observations of a random variable as a function of the time lags. For example, the lag-1 autocorrelation is the correlation between the current values of the time series and its values one time period before. The lag-2 autocorrelation is the correlation between the current values of the time series and its values two time periods before, and so on. A significant autocorrelation at a specific lag-k indicates that there is a repeating pattern in the data that occurs every (k) time periods. The ACF is typically plotted as a graph, with the lags on the x-axis and the autocorrelation coefficients on the y-axis. The confidence interval for the autocorrelation coefficients is also plotted, to help determine which coefficients are statistically significant. Bars exceeding the confidence interval may indicate that there is a significant correlation between the time series and itself at that lag.

The partial autocorrelation function (PACF) is a measure of the correlation between a time series and itself after removing the effects of the previous lags. In other words, it measures

the unique correlation between the time series and itself at a particular lag, without considering the correlations introduced by shorter lags in between. In this way, the influence of the lags between the current period and the specific lag under evaluation is removed. The PACF is calculated by regressing the time series values y_t on its own lagged values $y_{t-1}, \cdots, y_{t-k}$. At a specific lag-k, the PACF coefficient ($PACF_k$) is the coefficient of the last term α_k in the regression equation $y_t = \alpha_1 * y_{t-1} + \cdots + \alpha_k * y_{t-k} + e_t$. Like the ACF, the PACF is plotted as a graph, with the lags on the x-axis and the partial autocorrelation coefficients on the y-axis. The confidence interval for the partial autocorrelation coefficients is also plotted, to determine which partial autocorrelation coefficients are statistically significant. Bars that exceed the confidence interval indicate that there is a significant correlation between the time series and itself at that lag, while controlling for the effect of all the lags between the current value and the lag of interest. For example, if the PACF plot shows a spike at lag-k, this indicates that there is a significant correlation between the time series and itself at lag-k, after removing the influence of the shorter lags 1, ⋯ ,*k*-1. The autocorrelation function (ACF) at lag-k between current period data and lagged period data is given by the following formula [3]:

$$ACF_k(y_t, y_{t-k}) = \frac{\sigma^2(y_t, y_{t-k})}{\sigma^2(y_t)}$$

while the partial autocorrelation function (PACF) at lag-k is expressed as:

$$PACF_k(y_t, y_{t-k}) = \alpha_k$$
$$y_t = \alpha_1 * y_{t-1} + \cdots + \alpha_k * y_{t-k} + e_t$$

where α_k is the linear regression coefficient for the corresponding lag, and the linear regression equation has no intercept term.

In an MA(q) process the ACF has the property:

$$ACF_k \cong 0 \quad \forall k > q \ ,$$

that is the ACF of an MA(q) process cuts off after (q) lags, while the PACF decays to zero exponentially. For an AR process the ACF decays to zero exponentially, while the order can be determined by considering the PACF. An AR(p) process is characterized by the following property:

$$PACF_k \cong 0 \quad \forall k > p \ .$$

The PACF of an AR(p) process cuts off after (p) lags: The intuitive explanation is that for the AR(p) process y_t is expressed in terms of $y_{t-1}, \cdots, y_{t-p}$, so that additional lagged regressors y_{t-k} with k > p, have zero coefficients.

The ARIMA model parameters can be determined by means of visual inspection of the ACF and PACF plots. The interpretation of visual analysis is as follows: In the ACF plot, significant spikes that exceed the 95% statistical confidence interval, but do not appear in the PACF plot, suggest the presence of a moving average (MA) component. The order of the MA component can be estimated based on the highest significant lag in the ACF plot. If the ACF has (q) significant lags before dropping to zero and PACF has an exponential decay, we can identify the need for a moving average MA(q) order. Conversely, significant spikes in the PACF plot, but not present in the ACF plot, suggest the presence of an autoregressive (AR) component. The order of the AR component can be estimated based on the highest significant lag in the PACF plot. If the ACF has an exponential decay and the PACF has (p) significant lags before dropping to zero, we can identify the need for an autoregressive AR(p) order. When significant spikes are observed both in the ACF and PACF plots, it indicates the potential involvement of both AR and MA components. In such cases, exploring different model orders and assessing their goodness of fit becomes crucial. Model selection criteria like AIC or BIC can help determine the optimal model order.

In the previous section we determined that the time series that in the training range was first-order trend stationary was the differenced time series. For this time series the ACF is shown in Figure 1 and the PACF in Figure 2. The upper and lower limits for the confidence levels are given by: ±1.96/√*n*, where *n* is the number of observations in the training range, and the upper and lower limits are determined for 95% statistical confidence. The bars outside the confidence levels suggest that the corresponding lags are statistically significant. In the ACF plot in Figure 1, the first and second lags are statistically significant, suggesting there is the need for an MA(1) or MA(2) component in the ARIMA specification. Moreover, some spikes at higher lags are also statistically significant. However, in this case the MA order would still be 2, as the spikes at higher lags are likely due to noise or seasonality and should not be considered when estimating the MA order. For the PACF represented in Figure 2, the first and second lags are statistically significant, signaling the need for AR(1) or AR(2) coefficients within the ARIMA specification.

Figure 1. Autocorrelation function

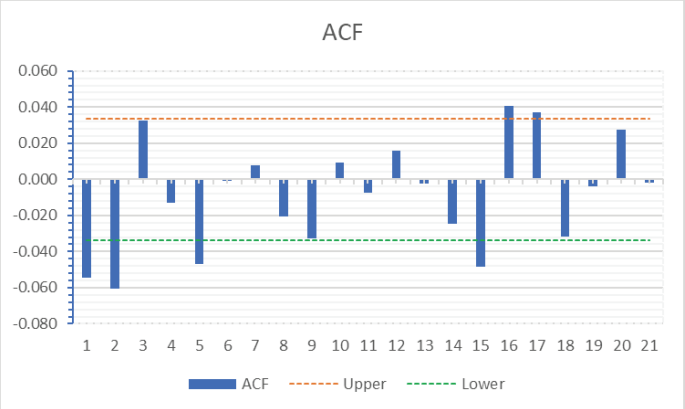
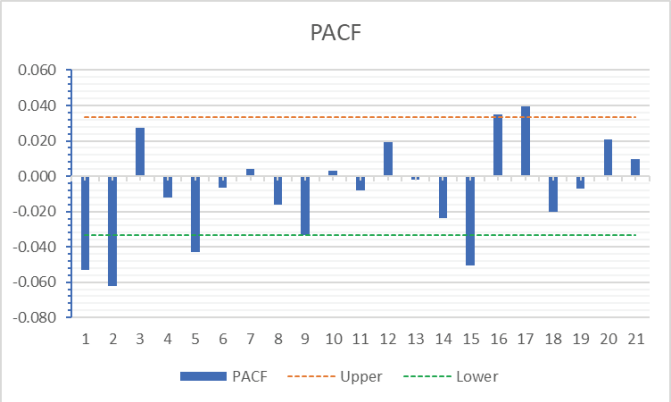


Figure 2. Partial autocorrelation function



In summary, the graphical representation of the autocorrelation function (ACF) with bars and the 95% confidence interval can provide insights into the presence of significant autocorrelations in a time series dataset. The graph can be interpreted as follows:

- *Bars inside the confidence interval:* If the bars representing the ACF values are mostly inside the 95% confidence interval, it suggests that the autocorrelations at those lags are likely due to random chance (null hypothesis). In other words, they are not statistically significant. This pattern indicates that the time series data does not have strong autocorrelations at those particular lags.
- *Bars outside the confidence interval:* If the bars representing the ACF values extend outside the 95% confidence interval, it suggests that the autocorrelations at those lags are likely not due to random chance and might be significant. This pattern indicates that the time series data has strong autocorrelations at those particular lags.
- *First few lags:* If the bars outside the confidence interval are primarily concentrated at the first few lags (lags 1 or 2), it might suggest the presence of a trend or a seasonal pattern.
- *Large lag values:* If there are bars outside the confidence interval at larger lag values, it might indicate the presence of noise or more complex patterns that extend over longer periods.

Based on the plots in Figure 1 and in Figure 2, AR(2) and MA(2) orders seem plausible given the significant spikes at lags-1 and lag-2 in the ACF and PACF plots.

2.5 SARIMA Model Specification

Seasonal Auto Regressive Integrated Moving Average (SARIMA) model specification consists of determining the order of the autoregressive (p), integration (d), moving average (q), seasonal autoregressive (P), seasonal integration (D), and seasonal moving average (Q) parameters. Integration or differencing (d) order is obtained when time series becomes first-order trend stationary. Seasonal integration or differencing (D) order is obtained when time series becomes first-order seasonal stationary. Autoregressive (p) and moving average (q) orders are obtained through autocorrelation and partial autocorrelation functions patterns and significant lags. Seasonal autoregressive (P) and seasonal moving average (Q) orders are obtained through seasonal autocorrelation and seasonal partial autocorrelation functions patterns and significant seasonal lags. The SARIMA general model is outlined in the Appendix. By customizing this model with the specific parameters, it is possible to derive the formulas of all the forecasting models presented in Section 3.

The Seasonal ACF (SACF) is an extension of the ACF that focuses specifically on the seasonal aspects of a time series. It measures the correlation between a time series and its past values at seasonal lags, by looking at how the data at a given point in time is related to its values at the same point in previous seasons. The SACF is represented as a plot with lags on the x-axis and the seasonal autocorrelation coefficients on the y-axis. Each bar on the SACF plot represents the correlation between the seasonal differenced time series and its values at a certain number of previous lags. By examining the SACF plot, we can identify the presence of seasonal patterns and determine the periodicity. Peaks in the SACF at specific lags correspond to the seasonality of the data. For example, if there are significant peaks at lags of 12 in monthly data, it suggests an annual seasonality, while peaks at lags of 21 in daily data suggest a monthly seasonality. The information from the SACF helps select appropriate SARIMA models, thus accounting for seasonal effects. The seasonal ACF and PACF in the training range are shown in Figure 3 and Figure 4 respectively. When the bars are either above or below the upper and lower bands of the 95% confidence interval, the corresponding lags are statistically significant.

Figure 3. Seasonal autocorrelation function

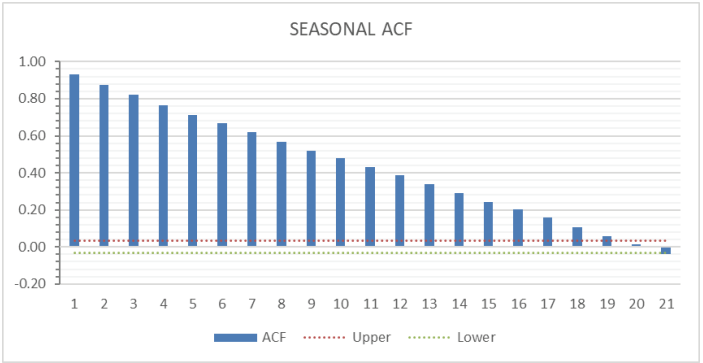
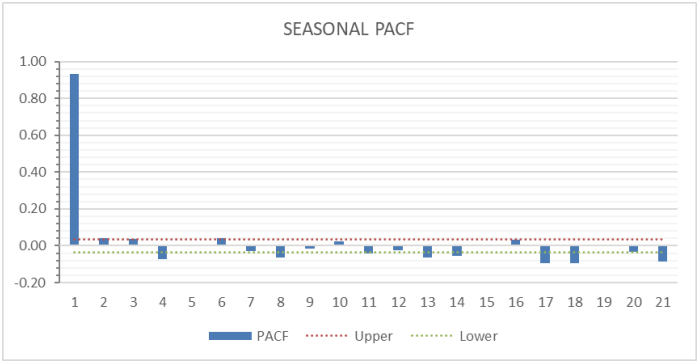


Figure 4. Seasonal partial autocorrelation function



Similar to the Seasonal Autocorrelation Function (SACF), which focuses on the correlation between the time series and its past values at seasonal lags, the Seasonal Partial Autocorrelation Function (SPACF) specifically quantifies the seasonal partial correlation while controlling for the influence of shorter lags. This helps identify the relationships related to seasonality. The SPACF is represented as a plot with lags on the x-axis and the seasonal partial autocorrelation coefficients on the y-axis. Each bar on the SPACF plot represents the partial autocorrelation between the current season's data and the data from the corresponding season in the past, while controlling for the effects of shorter lags. The SACF in Figure 3 shows an exponential decay, while the SPACF in Figure 4 shows that lag-1 is statistically significant, thus suggesting the need for a seasonal autoregressive AR model.

3. ARIMA Models Training and Testing

We train the ARIMA models considering as the objective function to minimize the sum of squared errors (i.e., SSE). In any of the ARIMA models with one or more moving average MA coefficients, the forecasting errors or residuals are not observed but they are estimated, therefore the model coefficients need to be calculated by solving an optimization problem to minimize the SSE objective function.

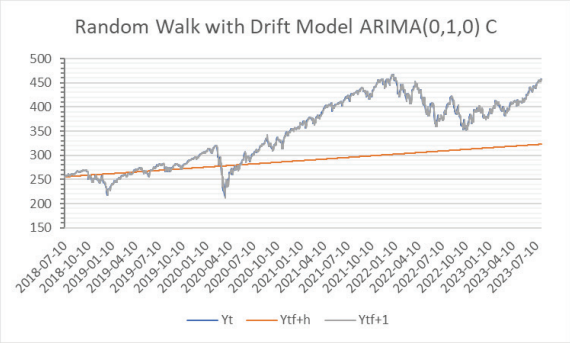
3.1 Random Walk with Drift Model

The Random Walk with Drift model consists of forecasting the current period data equal to the previous period data plus a constant drift α. In the ARIMA notation is expressed as: ARIMA(0,1,0)[C] with constant drift α, and the formula is:

$$\hat{y}_t = \alpha + y_{t-1}$$
$$initial\ \hat{y}_t \approx initial\ y_t.$$

In the training range we determine the α coefficient, while in the testing range we perform multistep forecasting for all the testing range, and one-step forecasting without re-estimation (the α parameter estimated in the training range remains constant in the testing range). The multistep forecast $\hat{y}_{t+h}$ for all the testing range and the one-step forecast without re-estimation $\hat{y}_{t+1}$ along with the actual data y_t are shown in Figure 5. The constant drift parameter α is computed with linear regression on the first-order differenced time series in the training range, and then it is kept constant in the testing range.

Figure 5. Random walk with drift in the testing range



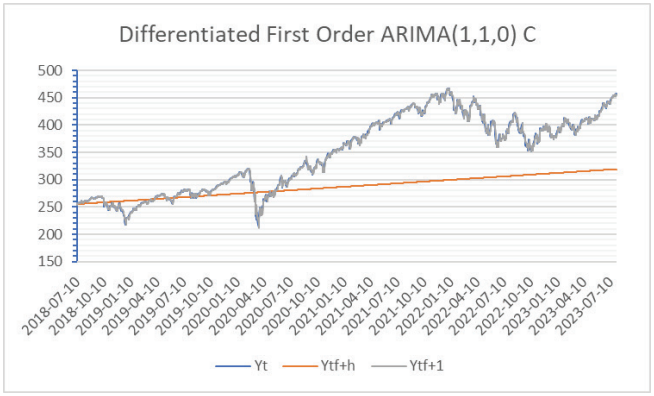
3.2 Differenced First-Order Autoregressive Model

The Differenced First-Order Autoregressive model consists of the difference forecast equal to the deterministic linear relationship with previous period data difference plus a constant drift. The ARIMA notation is: ARIMA(1,1,0)[C] with constant drift α, and the formula is given by:

$$\hat{y}_t = \alpha + y_{t-1} + \beta_1 * (y_{t-1} - y_{t-2})$$
$$initial\ \hat{y}_{t+i|i=0\rightarrow 1} \approx initial\ y_{t+i|i=0\rightarrow 1},$$

where the coefficients α and β_1 are computed by means of linear regression on the first-order differenced time series in the training range. In the testing range we perform multistep forecasting on the whole testing range and one-step forecasting without re-estimating the regression coefficients α and β_1 computed on the training range. The p-values of α and β_1 are < 0.05; hence, the coefficients are statistically significant with 95% statistical confidence. The results are shown in Figure 6.

Figure 6. Differenced first-order AR in the testing range



3.3 Brown Simple Exponential Smoothing Model

The Brown Simple Exponential Smoothing (SES) model consists of forecasting without trend or seasonal patterns. The ARIMA notation is: ARIMA(0,1,1) and the formula is:

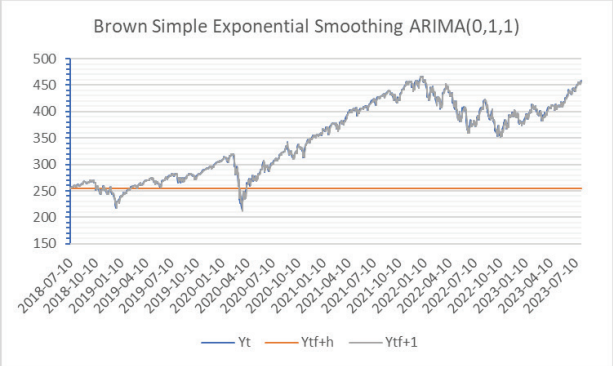
$$\hat{y}_t = y_{t-1} + \gamma_1 * e_{t-1}$$
$$initial\ \hat{y}_t \approx initial\ y_t,$$

where γ_1 is the moving average coefficient multiplied by the previous period forecasting error or residual. The forecasting errors or residuals are not observed but are estimated with the model, hence, the γ_1 coefficient is computed with the solver by minimizing the SSE.

Multistep forecasting within the testing range involves a process where data from the testing range is not used. The initial step within the testing range serves as the starting point, and from the second step onward, the forecast for each period is identical to the forecast for the previous period. This is because, from the second step onward, we no longer have forecasting errors or residuals to work with. These residuals are estimated and not directly observed, making it impossible

to calculate the second part of the formula, which involves the term $\gamma_1 * e_{t-1}$. Consequently, we can only calculate the forecast for the previous period, which remains fixed and is not updated with the corresponding data from the testing range. In contrast, in one-step forecasting without re-estimation, our forecast is continuously updated with actual data in the testing range, ensuring that it reflects the most recent information available. The results are shown in Figure 7.

Figure 7. Brown SES in the testing range



3.4 Simple Exponential Smoothing with Growth Model

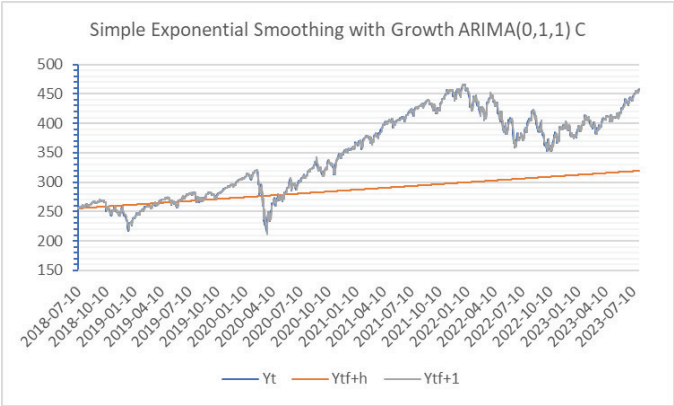
The Simple Exponential Smoothing with Growth model consists of adding a constant drift —to the Brown SES model. The ARIMA notation is: ARIMA(0,1,1)[C] with constant drift —. The equation is expressed as:

$$\hat{y}_t = \alpha + y_{t-1} + \gamma_1 * e_{t-1}$$
$$initial\ \hat{y}_t \approx initial\ y_t$$

The corresponding moving average coefficient is related to the previous period forecast error or residual. The corresponding forecast errors or residuals are not observed, but they are estimated, therefore the calculation of the coefficient $\gamma 1$ is done with the solver by minimizing the SSE.

In the context of multistep forecasting in the testing range, it is important to note that from the second step onwards, we lack the moving average component of the equation. Forecast errors for these steps are estimated, not observed. Thus, our calculations solely involve the constant drift from the training range, which is unaltered in the testing range. The multistep forecast is generated at the beginning of the testing range, covering the entire testing period. Importantly, no testing range data is incorporated; we rely on the previous period forecast. In one-step forecasting without re-estimation, we start with the initial testing range data. From the second step onward, we use the constant drift from the training phase, which remains fixed in testing. This drift is combined with the previous period data and the moving average coefficient multiplied by the previous period forecast error or residual. The results are shown in Figure 8.

Figure 8. SES with growth in the testing range



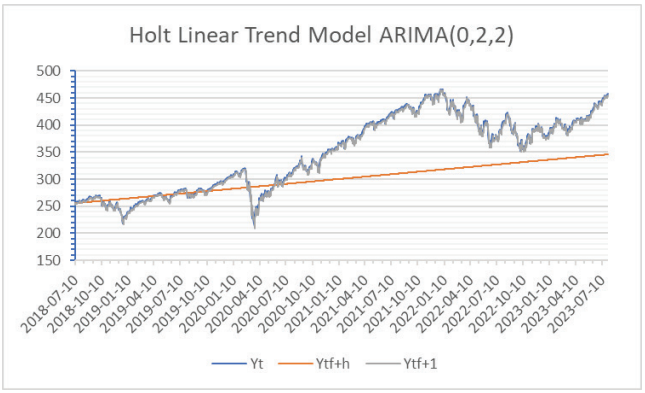
3.5 Holt Linear Trend Model

The Holt Linear Trend model is a forecasting technique designed to capture and predict time series with a linear trend. It is an extension of the SES method that accounts for both the level (the data's average) and the trend (the data's linear increase or decrease) components. The Holt Linear Trend model is useful for time series data that exhibit a consistent linear trend over time. The ARIMA notation is: ARIMA(0,2,2) and the equation is:

$$\hat{y}_t = 2 * y_{t-1} - y_{t-2} + \gamma_1 * e_{t-1} + \gamma_2 * e_{t-2}$$
$$initial\ \hat{y}_{t+i|i=0\rightarrow 1} \approx initial\ y_{t+i|i=0\rightarrow 1}$$

The model consists of two moving average coefficients, which are related to previous period forecast errors or residuals that are not observed but are estimated. Therefore, we need to perform the calculation using the solver to minimize the objective function SSE. The results are shown in Figure 9.

Figure 9. Holt linear trend model in the testing range



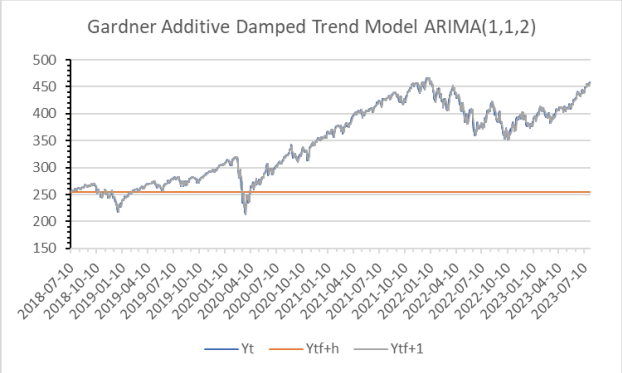
3.6 Gardner Additive Damped Trend Model

The Gardner Additive Damped trend model consists of forecasting with a damped linear trend pattern. It is a method used for forecasting data with a linear trend that is expected to dampen over time. The ARIMA notation is: ARIMA(1,1,2) and the formula is expressed as:

$$\hat{y}_t = y_{t-1} + \beta_1 * (y_{t-1} - y_{t-2}) + \gamma_1 * e_{t-1} + \gamma_2 * e_{t-2}$$
$$initial\ \hat{y}_{t+i|i=0\to 1} \approx initial\ y_{t+i|i=0\to 1}$$

The corresponding forecast errors or residuals are not observed but are estimated by minimizing the objective function SSE. The results are shown in Figure 10.

Figure 10. Gardner additive damped trend in the testing range



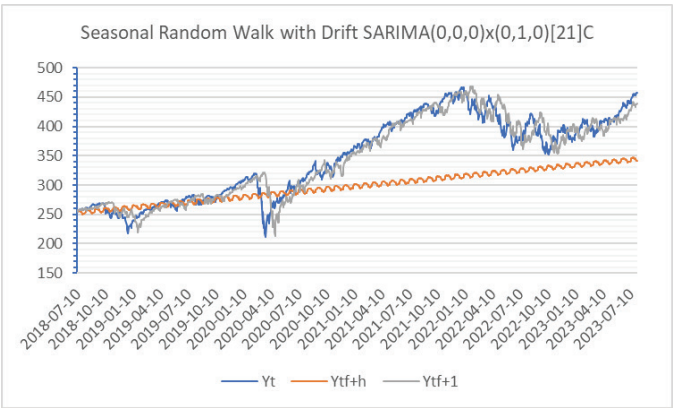
3.7 Seasonal Random Walk with Drift Model

The Seasonal Random Walk with Drift model consists of forecast equal to the previous season data plus a constant seasonal drift α. The seasonal ARIMA notation is: SARIMA(0,0,0)x(0,1,0)m[C] with constant drift α. In the non-seasonal part *m* corresponds to the forecasting seasonal lag. Since we are using daily data on a business calendar, the monthly seasonality is approximated as 21 days. The formula is expressed as:

$$\hat{y}_t = \alpha + y_{t-m}$$
$$initial\ \hat{y}_{t+i|i=0\to m-1} \approx initial\ y_{t+i|i=0\to m-1}$$

In this case we only have to estimate the constant seasonal drift coefficient based on seasonal differenced data by means of linear regression. The results are shown in Figure 11.

Figure 11. Seasonal random walk with drift in the testing range



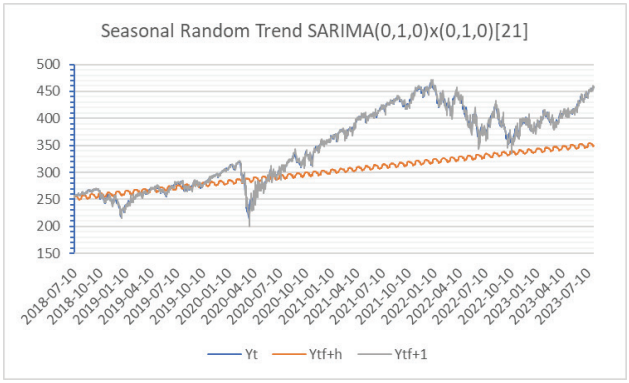
3.8 Seasonal Random Trend Model

The Seasonal Random Trend model consists of forecast equal to previous period data plus previous period seasonal difference. The seasonal ARIMA notation is: SARIMA(0,1,0)x(0,1,0)m without a constant drift. The formula is given by:

$$\hat{y}_t = y_{t-m} + y_{t-1} - y_{t-1-m}$$
$$initial\ \hat{y}_{t+i|i=0\to m} \approx initial\ y_{t+i|i=0\to m}.$$

The Seasonal Random Trend model does not require any parameter estimation, and the results are reported in Figure 12.

Figure 12. Seasonal random trend in the testing range



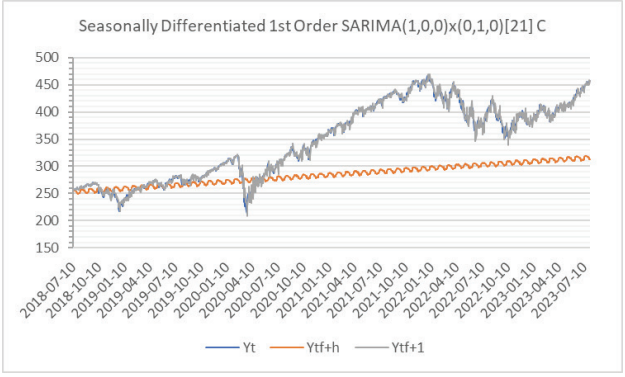
3.9 Seasonally Differenced First-Order AR Model

The Seasonally Differenced First-Order Autoregressive model consists of seasonal difference forecast equal to simple linear relationship with previous period seasonal difference data plus a constant seasonal drift —.The SARIMA model is: SARIMA(1,0,0)x(0,1,0)m[C] with constant drift —,and the formula is expressed by:

$$\hat{y}_t = \alpha + y_{t-m} + \beta_1 * (y_{t-1} - y_{t-1-m})$$
$$initial\ \hat{y}_{t+i|i=0\to m} \approx initial\ y_{t+i|i=0\to m}$$

The model coefficients are estimated through linear regression on the seasonally differenced time series. The results are shown in Figure 13.

Figure 13. Seasonally differenced 1st order AR in the testing range



3.10 ARIMA Model Selection and Forecasting Accuracy

This Section is dedicated to the selection of the ARIMA models described in Section 3, by using information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These criteria provide quantitative measures to assess the goodness of fit and complexity of ARIMA models. In this Section we want to assess which is the ARIMA model with the smallest value of the corresponding information criteria, hence, having the least information loss. The information criteria values are reported in Table 5.

Table 5. ARIMA model information criteria results

	ARIMA(0,1,0) C	ARIMA(1,1,0) C	ARIMA(0,1,1)	ARIMA(0,1,1) C	ARIMA(0,2,2)	ARIMA(1,1,2)	SARIMA(0,0,0) x(0,1,0)[21] C	SARIMA(0,1,0) x(0,1,0)[21]	SARIMA(1,0,0) x(0,1,0)[21] C
	Ytf	Ytf	Ytf	Ytf	Ytf	Ytf	Ytf	Ytf	Ytf
sse	5573.78003	5557.16771	5564.70877	5554.89410	5556.35688	5543.19346	81604.22480	11124.71125	10742.16702
aic	1672.39196	1664.17169	1666.81490	1660.77054	1661.67207	1653.55074	10861.74908	4038.70264	3920.88991
aicc	1672.39546	1664.17871	1666.81841	1660.77405	1661.67558	1653.55424	10861.75259	4038.70614	3920.89693
bic	1684.66908	1682.58739	1679.09203	1673.04767	1673.94920	1665.82787	10874.02621	4050.97976	3939.30561

The lowest values for both AIC and BIC correspond to the Gardner Additive Damped Trend ARIMA(1,1,2) model. Hence, this model has the least information loss. It is worth noting that BIC is more restrictive regarding the number of model parameters than AIC. The SARIMA forecasting models according to AIC and BIC have a higher information loss than the non-seasonal ARIMA models. If the non-seasonal ARIMA models have smaller AIC and BIC values compared to the SARIMA models, it suggests that the non-seasonal models provide a better trade-off between goodness of fit and complexity according to these information criteria. However, this conclusion should not be solely based on information criteria. While smaller AIC and BIC values generally indicate a better fitting model, it does not definitively prove the absence of seasonality in the time series. Seasonality might still be present, but perhaps not adequately captured by the SARIMA models.

We compared the forecasting accuracy metrics for all models in the testing range, both for multistep forecast and one-step forecast without re-estimation, where the model with the lowest RMSE and MAPE values has the highest forecasting accuracy. The numerical results are reported in Table 6 and in Table 7 for the multistep and one-step forecasting without re-estimation respectively.

Table 6. ARIMA multistep forecasting accuracy results

	ARIMA(0,1,0) C	ARIMA(1,1,0) C	ARIMA(0,1,1)	ARIMA(0,1,1) C	ARIMA(0,2,2)	ARIMA(1,1,2)	SARIMA(0,0,0) x(0,1,0)[21] C	SARIMA(0,1,0) x(0,1,0)[21]	SARIMA(1,0,0) x(0,1,0)[21] C
	Ytf+h	Ytf+h	Ytf+h	Ytf+h	Ytf+h	Ytf+h	Ytf+h	Ytf+h	Ytf+h
mae	61.79328	63.44402	93.44190	63.47141	51.83712	93.58275	53.15441	49.82970	65.53308
rmse	78.35324	80.26009	114.83651	80.28925	66.77482	114.96987	68.16878	64.36007	82.45431
mape	15.82517	16.24584	24.02651	16.25298	13.31903	24.06597	13.65761	12.82553	16.79963
mase	16.18418	16.61652	24.47322	16.62370	13.57658	24.51011	13.92159	13.05082	17.16367

Table 7. ARIMA one-step forecasting without re-estimation accuracy results

	ARIMA(0,1,0) C	ARIMA(1,1,0) C	ARIMA(0,1,1)	ARIMA(0,1,1) C	ARIMA(0,2,2)	ARIMA(1,1,2)	SARIMA(0,0,0) x(0,1,0)[21] C	SARIMA(0,1,0) x(0,1,0)[21]	SARIMA(1,0,0) x(0,1,0)[21] C
	Ytf+1	Ytf+1	Ytf+1	Ytf+1	Ytf+1	Ytf+1	Ytf+1	Ytf+1	Ytf+1
mae	2.99830	2.99562	3.00264	2.99790	3.72858	3.01025	13.66016	4.22343	4.17857
rmse	4.33688	4.32107	4.32304	4.32105	4.84383	4.34211	18.47327	5.99750	5.92746
mape	0.88514	0.88343	0.88561	0.88418	1.10882	0.88887	4.08202	1.24651	1.23579
mase	0.78528	0.78458	0.78642	0.78517	0.97655	0.78841	3.57771	1.10615	1.09440

For multistep forecasting the Seasonal Random Trend model SARIMA(0,1,0)x(0,1,0)[21] has the highest forecasting accuracy. In contrast, for one-step forecasting without re-estimation, the model with the lowest RMSE and MAPE values, i.e., the highest forecasting accuracy, is the Differenced First-Order Autoregressive model ARIMA(1,1,0)[C].

3.11 Residuals White Noise

A residual is a forecast error. It is the difference between the actual value of a time series and the predicted value. Residuals can be used to evaluate the accuracy of a forecasting model. If residuals are small, then the forecasting model is accurate. If residuals are large, then the forecasting model is inaccurate. Residuals can also be used to identify the sources of error in a forecasting model. For example, if the residuals are positively correlated, then this indicates that the forecasting model is underestimating the values of the time series. If the residuals are negatively correlated, then this indicates that the forecasting model is overestimating the values of the time series. Overall, residuals are a valuable tool for evaluating and improving the accuracy of forecasting models.

In this Section we analyze the requirement that forecasting errors or residuals should exhibit white noise behavior, so that they do not include any relevant information. A white noise is a stochastic process where each data point is *independently and identically distributed* (i.i.d.) with constant mean and constant variance. In simpler terms, white noise is a series of random data points that have no correlation with each other and with no evident pattern or trend over time. Although the terms "strong white noise" and "weak white noise" are not standard terminology in time series analysis or statistics, nevertheless, there is a concept related to the strength or intensity of white noise, and it typically refers to the behavior of the variance and the autocorrelation of the noise. In summary:

- *Strong White Noise:* Strong white noise typically refers to white noise with a constant variance (homoscedasticity) across all time points. In strong white noise, the variance of each data point remains the same throughout the time series. This means that the noise amplitude does not change over time. Properties are:
  - Observations are independent and identically distributed (i.i.d.);
  - Each observation is drawn from a normal (Gaussian) distribution with the same mean and variance;
  - There are no temporal dependencies or autocorrelations between observations at different time points.
- *Weak White Noise:* Weak white noise is often used to describe white noise with a time-varying variance (heteroskedasticity). In weak white noise, the noise variance can change from one time point to another. This implies that the noise amplitude can vary over time, potentially indicating some level of volatility or changing characteristics in the data. Properties are:
  - Observations are independent, but they may not be identically distributed;
  - There may be variations in the distribution (mean and/or variance) over time;
  - There are no temporal dependencies or autocorrelations between observations at different time points.

Hence, strong white noise is a stochastic or random process with constant parameters through time, which does not exhibit trend or seasonality and is statistically independent. It is required for model selection so that its residuals or forecast errors do not include any relevant information. A time series that is a white noise process exhibits several key properties, which are indicative of randomness and lack of systematic patterns. These properties include:

- *Independence:* Each observation in a white noise time series is independent of all other observations. This means that knowing the value of one data point does not provide any information about the values of other data points in the series;
- *Constant Mean:* The series has a constant and unchanging mean (average) value, which remains the same across all time periods;
- *Constant Variance:* The variance of the data points is constant and does not change with time;
- *Zero Autocorrelation:* There is no significant autocorrelation between observations at different time lags. In other words, the correlation between any two data points separated by a fixed number of time periods (lags) is approximately zero;
- *Gaussian (Normal) Distribution:* The data points in a white noise process typically follow a Gaussian (normal) distribution. This means that the data is symmetrically distributed around the mean, and most observations cluster near the mean, with fewer observations in the tails of the distribution;
- *Zero Cross-Correlations:* When comparing a white noise time series with another time series (cross-correlation), the cross-correlation function should be close to zero for all lags, indicating no systematic relationship between the two time series;
- *Randomness:* White noise is characterized by a lack of systematic patterns, trends, or seasonality. The data appears random, and there is no apparent structure in the time series;
- *Unpredictability:* Because of its randomness and lack of autocorrelation, white noise is essentially unpredictable. Forecasting future values of a white noise time series is challenging because there are no patterns to exploit.

Pure white noise is an idealized concept and is rarely encountered in real-world data. Real-world financial time series often contain some level of noise and randomness, but they also typically exhibit various patterns, trends, seasonality, and autocorrelation. White noise serves as a reference for assessing the randomness and independence of data points in a time series, but many time series will deviate from these ideal properties to varying degrees, while still containing valuable information. Although white noise is a theoretical idealization of randomness and unpredictability, nevertheless, understanding whether a time series exhibits white noise characteristics or not has its relevance in various situations:

- *Model Adequacy Testing*: In time series analysis, one of the fundamental tasks is to build predictive models for future values based on historical data. Checking for white noise properties in the forecasting errors or residuals helps to assess whether the model captures all relevant information (i.e., the signal) in the data, leaving only random noise in the residuals. If the residuals behave like white noise (i.e., are unpredictable and lack systematic patterns), then the model may be adequately capturing the signal, and any deviations from predictions might be due to random fluctuations;
- *Statistical Inference:* In many statistical methods, it is assumed that the residuals (the differences between observed and predicted values) follow a white noise process. Deviations from white noise behavior in the residuals may indicate that the model assumptions are violated, potentially leading to incorrect inferences;
- *Forecasting:* White noise time series are essentially unforecastable because they lack any visible patterns. If the time series resembles white noise, it implies that trying to forecast future values might be useless, and may need to consider alternative strategies for decision-making;
- *Noise Identification:* Separating the noise (random fluctuations) from the signal (systematic patterns) in a time series is crucial for understanding the underlying phenomena. Identifying and modeling the noise component can help isolate the true underlying trends and relationships in the data.

Not all time series are white noise processes, and many real-world time series contain meaningful patterns, trends, seasonality, and autocorrelation, which are essential to understanding and forecasting phenomena. The key issue is to assess whether the deviations from white noise behavior are due to meaningful patterns or if they are random fluctuations.

3.12 The Ljung-Box Test

The Ljung-Box test is a statistical test used in time series analysis to check whether a given time series is a white noise process or exhibits significant autocorrelation at various lags. The test consists of evaluating whether a time series is statistically independent with null hypothesis that data does not exhibit serial correlation and is statistically independent and independently distributed [8]. Ljung and Box introduced a statistical measure to assess the fitness of time series models. The measure, known as the Ljung-Box Q test statistic, is designed to evaluate the presence of autocorrelation in the residuals of a time series forecasting model. Autocorrelation refers to the correlation of a time series with its own past values at different lags. This test helps to determine whether the assumed independence of residuals holds true, and if not, it suggests that the model might need further refinement. The Ljung-Box test is based on the following steps:

- *Formulate the Null Hypothesis:* The null hypothesis is that there is no autocorrelation in the time series at any of the selected lags;
- *Choose a Significance Level (alpha):* Typically, a significance level, denoted as alpha, is chosen to determine the threshold for statistical significance. In this work alpha is 0.05;
- *Select the Lag Values (h):* We need to specify the number of lags to test for autocorrelation. The choice of lags often depends on the specific application and the nature of the data. In this work the number of lags is 21, corresponding to a monthly seasonality;
- *Calculate the Test Statistic: T*he Ljung-Box Q test statistic is calculated with the following formula:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}(e_t, e_{t-k})^2}{n-k},$$

where n is the number of observations and k, the corresponding lag within the sum, varies from the first to the last of the h lags for which we are performing the Ljung-Box test;

- *Compare the Q Test Statistic to the Chi-Square* (χ^2) Distribution: Compare the calculated Q test statistic to the χ^2 distribution with degrees of freedom equal to k. The degrees of freedom are typically equal to the number of lags being tested (k). The critical value for the chi-square distribution is computed for the chosen significance level;
- *Make a Decision:* Deciding based on the Q test statistic involves two approaches:
  - *Using Critical Values*: If the calculated Q test statistic exceeds the critical value derived from the chi-square distribution, the null hypothesis is rejected, indicating significant autocorrelation in the time series at one or more lags. Conversely, if the Q test statistic is less than the critical value, the null hypothesis cannot be rejected, suggesting no significant autocorrelation;
  - *Using p-values:* Alternatively, we can assess the Q test statistic's p-value. If the p-value for h lags exceeds 0.05, it means that none of the h lags have significant autocorrelation. In such case, the time series of forecasting errors or residuals is independently distributed, and the null hypothesis cannot be rejected with 95% statistical confidence. Conversely, if the p-value is less than or equal to the chosen significance level alpha, the null hypothesis is rejected, indicating significant autocorrelation in the time series at one or more lags.

We consider the ARIMA(1,1,0)[C] forecasting model, a differenced first-order autoregressive model, which obtained the highest forecasting accuracy in the testing range for one-step forecasting without re-estimation, as shown in Table 7. The ACF and PACF of the forecast errors or residuals for the ARIMA(1,1,0)[C] model are represented in Figure 14 and Figure 15 respectively.

Figure 14. ARIMA(1,1,0)[C] residuals ACF

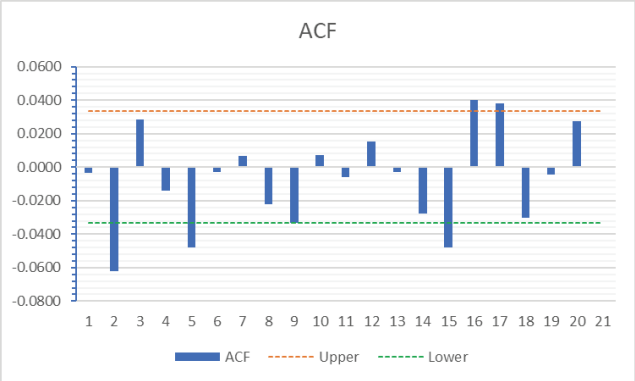
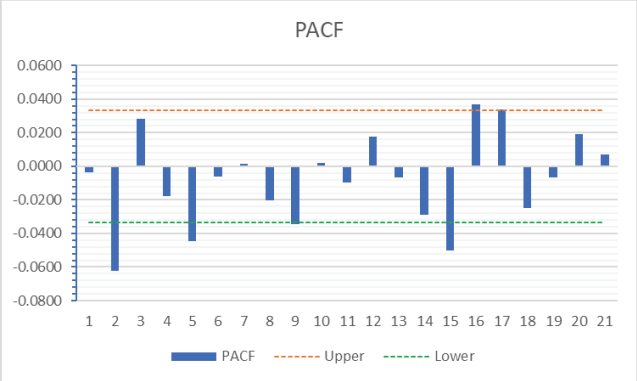


Figure 15. ARIMA(1,1,0)[C] residuals PACF



We assessed the individual statistical significance of forecast errors or residuals by examining the ACF plot to identify lags outside the 95% confidence range. If lags fall outside the confidence interval, the forecast errors or residuals for those periods are considered individually statistically significant. The ACF results are corroborated by the PACF, and any lags outside (either above or below) indicate their individual statistical significance. Given the results of the Q test statistic with p-values < 0.05 shown in Table 8, we can conclude that there is group lag autocorrelation or serial autocorrelation for some of the 21 lags. This observation aligns with what we observe in both the ACF and PACF graphs, where lags outside the 95% confidence range indicate individual statistical significance. Consequently, the corresponding residuals for these lags contain relevant information that should be considered and incorporated into the model.

If the Ljung-Box p-values are < 0.05 for all the lags and there are lags outside the 95% confidence intervals in the ACF plot, it suggests that the time series of forecast errors or residuals is not white noise and there is significant autocorrelation in the residuals. The fact that there are some lags within the 95% confidence interval in the ACF plot does not necessarily mean that the residuals are white noise. It is common for some lags to fall within the confidence interval even if there is some autocorrelation present. However, the presence of lags outside the confidence interval indicates that the autocorrelation is significant. It is possible for the ACF value to be within the 95% statistical confidence interval while the Ljung-Box test p-value is < 0.05 for a certain lag. The ACF plot measures the correlation between the residuals at different lags, while the Ljung-Box test measures the overall autocorrelation in the residuals up to a certain lag. Therefore, it is possible for the ACF value to be within the confidence interval for a specific lag, indicating that the correlation at that lag is not statistically significant, while the Ljung-Box test p-value is < 0.05, indicating that there is significant autocorrelation in the residuals overall and they are not white noise. The Ljung-Box test assesses the joint significance of autocorrelations at multiple lags simultaneously. If the p-value is < 0.05, it indicates that there is significant autocorrelation among a group of lags, even if individual ACF values may not be significant.

In summary, the Ljung-Box test considers the overall pattern of autocorrelation across multiple lags and assesses whether this pattern is statistically significant. This means that while individual ACF values for specific lags may not be significant, the combination of autocorrelations across several lags could still be statistically significant, leading to a Ljung-Box test p-value < 0.05.

The presence of significant autocorrelation detected by the Ljung-Box test, even when individual ACF values are within the confidence interval, indicates that there is autocorrelation in the overall pattern of the residuals, emphasizing the importance of examining the joint behavior of autocorrelations across multiple lags in time series analysis.

Therefore, the Ljung-Box test p-values reported in Table 8 suggest increasing the model complexity by adding more moving average terms. Consider the ARIMA(1,1,2) model described in Section 3.6. The Ljung-Box test p-values are reported in Table 9.

Table 8. ARIMA(1,1,0)[C] Ljung-Box Q test statistic p-values

Lag	1	2	3	4	5	6	7	8	9	10	
P-Value	0.83881	0.00132	0.00109	0.00214	0.00017	0.00040	0.00083	0.00088	0.00039	0.00072	
Lag	11	12	13	14	15	16	17	18	19	20	21
P-Value	0.00127	0.00170	0.00290	0.00199	0.00023	0.00006	0.00002	0.00001	0.00002	0.00001	0.00002

Table 9. ARIMA(1,1,2) Ljung-Box Q test statistic p-values

Lag	1	2	3	4	5	6	7	8	9	10	
P-Value	0.88863	0.98983	0.95403	0.96115	0.11907	0.18468	0.26617	0.24489	0.11116	0.15142	
Lag	11	12	13	14	15	16	17	18	19	20	21
P-Value	0.19552	0.21256	0.26530	0.21327	0.04843	0.01819	0.00802	0.00630	0.00935	0.00704	0.01030

It is possible to observe that after incorporating residuals into the ARIMA model, the number of lags where the residuals or forecast errors are not statistically significant, hence they do not exhibit statistically significant autocorrelation, has increased from 1 (as reported in Table 8) to 14 as reported in Table 9. The results of the Ljung-Box test for the ARIMA(1,1,2) residuals shown in Table 9 are confirmed by the ACF and PACF plots depicted in Figure 16 and Figure 17 respectively.

Figure 16. ARIMA(1,1,2) residuals ACF

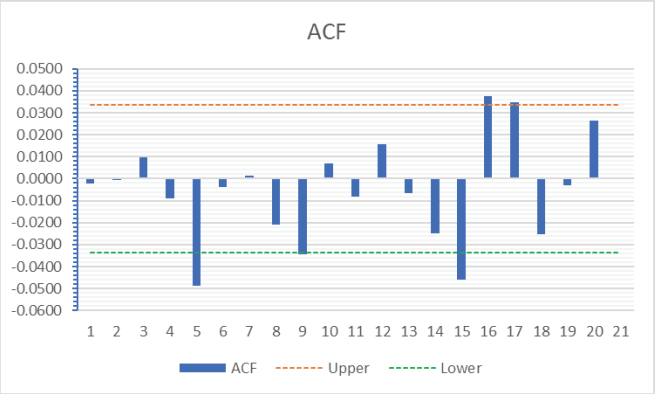
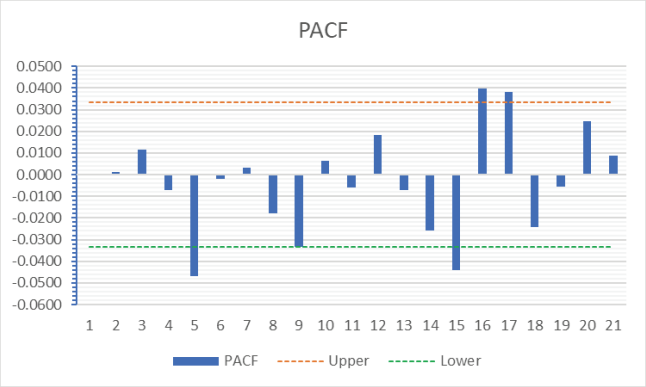


Figure 17. ARIMA(1,1,2) residuals PACF



4. Conclusions

Forecasting the stock market is a highly complex and dynamic task requiring a profound understanding of economic and financial factors. Both statistical methods and data-driven approaches are employed in financial forecasting, but the selection of the optimal model depends on various factors such as data availability, quality, and the specific characteristics of the analyzed data. Frequently, the available data is limited to the historical time series of the financial asset under analysis. Furthermore, it is mandatory to rigorously assess all the statistical hypotheses in order to understand the limitations and pitfalls of the selected model and to accurately interpret the results of the analyses.

This article has presented the third part of a series dedicated to the rigorous evaluation of statistical forecasting methods for financial time series. While the first part [1] established the general evaluation framework, the historical dataset, the accuracy metrics, and the simple forecasting baselines, and the second part [2] provided a comprehensive analysis of the exponential smoothing methods, this part has been dedicated to the AutoRegressive Integrated Moving Average models and to their seasonal extension. Beyond their forecasting capabilities, the ARIMA models have offered a unifying perspective on the series: several of the simple and exponential smoothing methods analyzed in the previous parts have been derived as particular cases of the general ARIMA template, thus connecting the models presented so far within a single, coherent formal framework.

The methodological path followed in this article illustrates how each modeling decision is motivated by the rigorous verification of the underlying statistical hypotheses. The assessment of first-order stationarity, through the deterministic trend and Phillips-Perron tests, guided the specification of the differencing order, while the deterministic seasonality test addressed the seasonal components of the time series. The autocorrelation and partial autocorrelation functions supported the identification of the autoregressive and moving average orders, and the model parameters were optimally estimated within the training range. The candidate models were then compared through the Akaike and Bayesian information criteria, and their forecasting accuracy was assessed in the independent testing range, considering both multistep forecasting and one-step forecasting without re-estimation. Finally, the Ljung-Box test was applied to the residuals of the selected models to verify their white noise property.

The experimental results highlight the importance of evaluating the models under different forecasting horizons and with complementary criteria. Within the training range, the lowest values of both the Akaike and Bayesian information criteria correspond to the Gardner Additive Damped Trend model, i.e., the ARIMA(1,1,2) model, which therefore exhibits the least information loss, while the seasonal models report a higher information loss than the non-seasonal ones. In the testing range, the Seasonal Random Trend model SARIMA(0,1,0) x(0,1,0)[21] obtained the highest forecasting accuracy for multistep forecasting, whereas the Differenced First-Order Autoregressive model ARIMA(1,1,0) with constant drift obtained the highest forecasting accuracy for one-step

forecasting without re-estimation. The models with the least information loss in the training range are therefore not necessarily those with the highest forecasting accuracy in the testing range, which confirms the importance of a validation process performed on data that were not used for the parameter estimation.

The residuals analysis completes the assessment and, at the same time, delineates the boundaries of the ARIMA framework. The Ljung-Box test applied to the residuals of the ARIMA(1,1,0) model with constant drift reported statistically significant autocorrelation for almost all the considered lags, indicating that the model does not capture all the information contained in the historical data. Increasing the model complexity by adding moving average terms improved this behavior: for the ARIMA(1,1,2) model, the number of lags without statistically significant autocorrelation increased from 1 to 14. Nevertheless, the residuals cannot yet be considered a white noise process, and this result is not accidental: ARIMA models assume a linear relationship and a constant variance of the forecast errors, an assumption that is frequently violated by financial time series, where periods of high and low volatility tend to cluster.

The contribution of this article lies not only in the analysis of the mathematical formulations and parameter optimization of the ARIMA models, but more importantly, in delineating a rigorous statistical methodology for assessing and utilizing these models. We introduced the concepts of first-order and seasonal stationarity, we derived the model specification from the autocorrelation analysis, and we evaluated the goodness of fit through the residuals white noise testing, to ensure that the selected models are aligned with the property of residuals randomness. Additionally, we employed reliable information loss criteria and forecasting accuracy metrics for a statistically robust model selection and validation process, performed within the training range and verified in the independent testing range.

Importantly, our findings highlight the adaptability of the ARIMA framework to the temporal dependencies of financial data, enabling short-term predictions without surrendering to overfitting, and with the interpretability that is essential for informed financial decision-making. While acknowledging the advantages of data-driven approaches in handling large datasets, we contend that the transparency and the statistical discipline offered by these methods remain paramount in the financial forecasting landscape, where clarity in the forecasting process is essential for informed decision-making.

The limitations revealed by the residual diagnostics naturally introduce the fourth and final part of this series, already planned for the IFTA Journal 2028, which will be dedicated to the advanced forecasting models. The ARIMA models will be extended by incorporating the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) equations, allowing the simultaneous modeling of both the conditional mean and the conditional variance of the time series, and thus the explicit representation of the volatility clustering observed in financial markets. The hybrid ARIMA-GARCH framework will be further extended with the threshold GJR-GARCH models, which account for the asymmetric impact of positive and negative shocks on volatility, and finally with the non-Gaussian GARCH models based on the Student's t distribution, which remove the dependency on the assumption of normally distributed returns and better capture the fat tails and excess kurtosis that frequently characterize financial data.

Appendix

$$\Delta^d \Delta_m^D y_t = \alpha + \sum_{i=1}^{p} \beta_i \Delta^d \Delta_m^D y_{t-i} + \sum_{j=1}^{q} \gamma_j e_{t-j} + \sum_{k=1}^{P} \omega_k \Delta^d \Delta_m^D y_{t-km} + \sum_{l=1}^{Q} \theta_l e_{t-lm} - \sum_{i=1}^{p} \sum_{k=1}^{P} \beta_i \omega_k \Delta^d \Delta_m^D y_{t-i-km} + \sum_{j=1}^{q} \sum_{l=1}^{Q} \gamma_j \theta_l e_{t-j-lm} + e_t$$

SARIMA model specification

References

[1] Pandini, D. "Prediction is very difficult: especially if it's about the financial markets! (Part I)," IFTA Journal 25, pp. 16-25, 2025.

[2] Pandini, D. "Prediction is very difficult: especially if it's about the financial markets! (Part II)," IFTA Journal 26, pp. 26-38, 2026.

[3] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. "Time Series Analysis: Forecasting and Control." 3rd Edition, Englewood Cliffs, NJ: Prentice Hall, 1994.

[4] Akaike, H. "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, 19(6), pp. 716-723, 1974.

[5] Schwarz, G. E. "Estimating the Dimension of a Model," Annals of Statistics, 6(2), pp. 461-464, 1978.

[6] Box, G. E. P., and Jenkins, G. M. "Time Series Analysis: Forecasting and Control." Holden-Day, San Francisco, CA, 1970.

[7] Phillips, P. C. B., and Perron, P. "Testing for a Unit Root in Time Series Regression," Biometrika, 75(2), pp. 335-346, 1988.

[8] Ljung, G. M., and Box, G. E. P. "On a Measure of Lack of Fit in Time Series Models," Biometrika, 65(2), pp. 297-303, 1978.

Footnotes

1 Master of Financial Technical Analysis (International Federation of Technical Analysts)

2 Certified Financial Technician (International Federation of Technical Analysts)

3 Chartered Market Technician (CMT Association)

4 Certified SIAT (Società Italiana Analisi Tecnica) Technical Analyst

The Reliability of MACD and RSI Crossovers During Geopolitical and Economic Crises Events: A Multi-Asset Technical Study

Sara Alyasiri, MFTA, CFTe
Amman, Jordan
sarahalyasiri2@gmail.com

Abstract

Financial markets frequently experience increasing volatility during economic crises, geopolitical tensions, and financial instability. During such environments, traders and analysts often rely on technical indicators to interpret price momentum and identify potential trading opportunities. This study evaluates the effectiveness of two widely used indicators the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI) during the time of four major crisis in the modern economy including the Global Financial Crisis (2008–2009), the COVID-19 market shock (2020), the inflation and geopolitical volatility of 2022, and the banking sector crisis in 2023.

Using daily price data for gold, the S&P 500, the Nasdaq-100, and Brent crude oil, trading signals were identified based on standard MACD crossovers and RSI overbought and oversold thresholds. A ten-day observation window was used to evaluate whether price movements aligned with the signals. The results show that indicator performance varies across assets and market conditions, with MACD generating more frequent signals while RSI produced fewer but more selective signals during periods of extreme volatility. Thus, the findings highlight the importance of adapting technical analysis strategies to changing market environments.

INTRODUCTION

1.1 Background of the Study

The stock market is a part of the multidimensional capital market, that is influenced by a multiple factor including the economic conditions, company performance, and market sentiment, Investors tend make decisions to buy or sell stocks based on the reasons that actually cause stock values to change, which makes it challenging for investors to analyze the volatile fluctuations in the stock prices. therefore; technical expertise, such as technical analysis, is essential to comprehend and predict stock price changes and within technical analysis, there is a concept of price trends, which refers to the patterns of price movements over a given period, such as upward trends (bullish), downward trends (bearish), or sideways trends. (Melda and Mira 2025)

Technical analysis is a method commonly used by investors and traders as an essential tool in trading activities in order to predict price direction, establish movement limits under certain conditions, and indicate target directions along with the associated risks that it carries, and it also serves in identifying the best time to buy and sell stocks in order to maximize capital gains. Technical analysis uses previous trading data for determining future price changes in financial products. Momentum indicators like the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) are widely used in technical analysis tools In technical analysis, these two indicators are frequently used. (Investopedia 2024)

During period of crisis the investor sentiment may rapidly shift, risk premiums are more likely lead to an increase, and capital flows often move unexpectedly between asset classes. therefore; the financial markets become more sensitive to new information, resulting in unexpected price movements that may not necessarily follow standard valuation frameworks. what happens in these circumstances specifically is that the market participants which are mainly the traders will usually rely on the technical analysis to analyze the price action and determine their potential trading opportunities at that time.

Two of the most common momentum indicators in technical analysis are the Relative Strength Index (RSI) and the Moving Average Convergence Divergence (MACD). Both pf these methods have been created to support the traders in determining the potential momentum fluctuations or shifts, it also supports setting the entry or exit signals, and trend reverses and corrections. Given that MACD analyzes the correlation between the moving averages to identify shifts in market momentum, where on the other hand the RSI assesses the speed and magnitude of price movements to identify overbought and oversold circumstances. (Singh and Priyanka 2025)

Aside of their wide popularity, the accuracy of these indicators during times of high market stress remains a significant concern in all times. Financial crises and geopolitical conflicts that occur worldwide can result into a significantly elevated levels of volatility, risk repricing, and panic-driven trading behavior that is a result of the sentimental changes in the trader's behavior, any reaction that is based on emotions or reactions will also impact the market. Under such form of circumstances, the price dynamics may in a way be different from the standard and the normal market patterns, therefore this might be resulting in reducing the efficiency of common technical indicators or in some cases it could be leading into generating a delayed and inaccurate trading signals.

Moreover; recent market developments dynamics have pointed out such concerns that included as mentioned the geopolitical concerns, tariff-driven trade wars, and uncertain macroeconomic conditions around the world all these factors have had a significant role in resulting in a form of high levels of market volatility by the year 2023, in additionally even during a rough and challenging macroeconomically environment, significant key assets in the market such as gold, S&P 500, and the Nasdaq have all succeeded to reach historical records. These developments have generated significant concerns about whether those technical indicators that are commonly used can still deliver real time and accurate warnings during times of geological instability and economic uncertainty.

(Moodley 2026)

The aim of this research paper is to study and evaluate the accuracy of the two popular momentum indicators the Relative Strength Index (RSI) and the Moving Average Convergence Divergence (MACD) perform during circumstances and conditions of high market volatility triggered on by significant geopolitical and economic crises. Furthermore, this study studies whether crossover signals generated through the RSI and MACD indicators can be still predictably reliable in unstable market conditions or whether their performance becomes less accurate as there is a form of instability and systemic risk rise.

In order to address this concern, the study considers indicator signals from major significant types of assets, which include gold, Brent crude oil, the S&P 500 index, and the Nasdaq Composite index, noting that these assets were specifically selected because of their importance on a global scale, liquidity, and sensitivity to the macroeconomic changes and geopolitical changes. Although crude oil prices are extremely sensitive to the geopolitical tensions, supply interruptions, and worldwide economic activity, the gold has historically been a safe-haven asset during all the difficult times. Moreover, the Nasdaq Composite and S&P 500 represent a crucial benchmark for international equity markets and can demonstrate the investor's sentiment and the tolerance for risks of these investors.

Moreover; as mentioned this study focus on several important phases of financial market crisis in the recent past two decades. These include the 2008-2009 Global Financial Crisis, the market disruption caused by the COVID-19 pandemic in 2020, the inflation and geopolitical volatility associated with the Russia-Ukraine conflict in 2022, and the banking sector stress episode observed in 2023 as a result of Silicon Valley Bank's collapse and instability in several financial institutions. These periods illustrate various forms of economic and financial shock that had a substantial impact on global financial markets. (Zakry and others 2026)

Now, by aiming to compare the performance of the RSI and MACD signals performance of RSI and MACD signals during times of crisis to their performance during a more stable market conditions in order to discover whether traditional technical indicators remain dependable tools during extreme market instability. The findings aim to contribute to a better understanding and also becoming more acquainted with of technical analysis in unstable market conditions, while also providing practical recommendations that can assist traders, analysts, and institutional investors improve decision-making and risk management strategies throughout times of macroeconomic uncertainty.

1.2 Research Questions

1. To what extent do the MACD and RSI, maintain their predictive accuracy during periods of extreme market volatility caused by major economic and geopolitical crises?
2. Does the effectiveness of these indicators vary across different asset classes and types of market shocks?
3. Can we rely on the MACD and RSI on daily time frame signals to trade during market volatility.

MATERIALS AND METHODS

This study implements a quantitative approach that evaluates the accuracy of frequently applied technical indicators during periods of increased market volatility that is a consequence of the major economical and geopolitical crises that occur in the world the past two decades. This research paper focuses on determining and measuring the dependability of momentum-based technical signals generated by the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) across the global financial crises that has impacted the world and the financial market that have occurred between the year 2008 and 2023. (TradingView 2026)

2.1 Data Selection and Asset Classes

The analysis mainly conducted on daily price data for four significant financial assets, including both commodity and equities markets. Gold, Brent crude oil, the s&p 500 index, and the Nasdaq Composite index were among the selected assets. These assets were chosen for their global significance, high liquidity, and susceptibility to macroeconomic developments and geopolitical events.

Among ages, the gold is commonly classified as an safe-haven asset, which leads to an increase of interest from shareholders to during times of uncertainty and financial stress. Brent crude oil is a significant benchmark for global energy markets, and it is particularly sensitive to geopolitical tensions, supply interruptions, and changes in global economic activity. Meanwhile, the S&P 500 and Nasdaq Composite indices are two of the most important global equity market benchmarks, signaling broader investor sentiment, technical sector performance, and general risk appetite.

Therefore; the daily pricing data allows an accurate evaluation that is more detailed of short-term market reactions to technical signals while being consistent throughout several crisis periods.

2.2 Technical Indicators

Two of the two popular momentum indicators: the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI) are mainly examined in this study and these indicators were selected because traders and analysts frequently rely on them to identify and address the potential trend reversals, momentum shifts, and entry or exit opportunities.

In order to determine on whether the market is overbought or oversold, the RSI is calculated using the standard 14-period setup. It measures the speed and magnitude of price changes of price movement and fluctuations. Overbought conditions are typically indicated by values above 70, whereas on the other hand oversold conditions are indicated by values below 30.

The MACD indicator is calculated using the commonly applied parameters of 12-period and 26-period exponential moving averages, with a 9-period signal line. The MACD measures the relationship between short-term and long-term momentum and generates trading signals when the MACD line crosses above or below the signal line.

2.3 Signal Identification

Trading signals are identified according to the commonly used crossover rules associated with each indicator.

When the RSI indicator crosses over 30, it provides a positive signal, which mainly tend to indicate a possible move from oversold to upward momentum, whereas a negative signal is issued when the RSI falls below 70, this could be a signal that

overbought conditions may be reversing. On the other hand the MACD indicator a bullish signal occurs when the indicator MACD line crosses over the signal line, which is indicating an increasing upward momentum, moreover; as the MACD line crosses below the signal line, it indicates that momentum is decreasing and that prices are likely to fall.

These signal definitions reflect frequently used technical analysis standards and provide the investor to consistent signal identification throughout market conditions.

2.4 Crisis Period Selection

To evaluate indicator performance under different market conditions, the study examines four major periods of financial market stress: the Global Financial Crisis (2008–2009), the COVID-19 pandemic shock (2020), the inflation and geopolitical volatility associated with the Russia–Ukraine conflict in 2022, and the banking sector crisis episode observed in 2023. These events were selected due to their substantial impact on investor sentiment, market liquidity, and price volatility across both commodity and equity markets. (Cheema, Faff, and Szulczyk 2023)

2.5 Signal Evaluation Method

The study evaluates the subsequent price performance following each RSI and MACD signal to assess their accuracy and efficiency. The methodology specifically examines at price movement during a 10-day trading timeframe following the generation of each signal.

A signal is considered to be successful if the asset price moves in the expected direction during the 10-day time frame. Bullish signals are classified as successful if they result in a positive price movement after the signal, whereas bearish signals are considered successful if prices fall throughout the evaluation time.

The 10-day evaluation is selected in order to capture short-term momentum following the technical signals while reducing the influence of short-term market fluctuation, because the momentum indicators such as RSI and MACD typically require several trading sessions for signals to fully develop, especially in a highly volatile environments whereas a more short evaluation window may therefore fail to capture the full impact of the signal, meanwhile a long horizons may introduce an additional market changes that has nothing to do with the original indicator signal.

Therefore, a specific 10-trading-day period would perhaps provide a balanced framework for measuring signal accuracy and effectiveness as it provides a sufficient time duration for momentum shifts to take place meanwhile also reduce the impact of irrelevant macroeconomic developments or additional market shocks that could change and impact the appraisal of the original technical signal so Tas mentioned the 10 trading day period time frame also contributes to consistent results when adjusting to compare indicator performance across assets and crisis periods.

Along with evaluating the price performance after each and every signal, the study evaluates how both the frequency and also the responsiveness of RSI and MACD signals fluctuation and changes during times of an increased volatility, in fact, the financial crises usually result in rapid fluctuations in prices and increased market noise and this leads to causing momentum indicators to generate signals more frequently and in other cases it provide contradicting signals in short periods of time. Observing these patterns provides more information that must be studied when evaluating the effectiveness of technical indicators during extreme market conditions.

The full dataset of trading signals used in the analysis is presented in Appendix A.

RESULTS

3.1 Global Financial Crisis (2008–2009)

This section demonstrates the empirical data from studying the effectiveness of the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI) during the Global Financial Crisis of 2008-2009. The investigation focused on four important financial assets representing various asset classes: gold, the S&P 500 index, the Nasdaq-100 index, and Brent crude oil.

Using daily price data, the trading signals generated by both indicators were discovered using the principles outlined in the methodology section. For each signal, the market price at the signal date was compared to the closing price ten trading days later to assess the indicator's predictive ability. However a bullish signal was considered successful if the price have increased after ten trading days, perhaps a bearish signal was also considered successful if the price have declined during the same time period. Therefore; The following subsections would summarize the studied and observed signals across the analyzed assets.

3.1.2 Evaluating the Performance of MACD and RSI Indicators on Gold During the Global Financial Crisis

During the Global Financial Crisis that have started in 2008, the gold market have experienced back then a significant volatility in which the investors have increasingly turned to the gold that is known to be the safe-haven asset among the widespread financial instability. Amd its important to highlight that MACD indicator have then generated 20 trading signals between both years 2008 and 2009 but many of these signals captured a form of a short-term directional movements among the gold market as prices have witnessed fluctuation during that time period of uncertainty. Also several bullish signals have taken place during the phases of recovery that were following temporary market corrections noting that bearish signals were often generated during short-term pullbacks within broader upward movements.

And during the same time frame, the RSI indicator gave out seven indications, because of the indicator's design, which requires extreme overbought or oversold conditions to generate such indications, in short, the RSI indicator have produced less alerts than MACD. However, some RSI indications correlated with significant turning points in the gold market, in particular during times of increased demand for safe-haven assets.

3.1.3 Evaluating the Performance of MACD and RSI Indicators on S&P 500 Index During the Global Financial Crisis

During the Global Financial Crisis, the S&P 500 index witnessed one of the greatest drops in recent financial history, factors as extreme volatility, significant price drops, and sudden recovery periods have occurred in the market collapse.

During the crisis, the MACD indicator have generated 23 signals. While some of these signals emphasized the challenge of interpreting momentum indicators during sudden market shocks, others were successful in identifying short-term trend reversals. At times, particularly during those in which we call extremely volatile circumstances, MACD signals appeared after significant price changes had already occurred due to quick price swings.

But during the exact time period, the RSI indicator gave five signals. Because RSI requires extreme market conditions in order to generate indications, it generated considerably fewer signals than the indicator MACD. Yet several of these indications appeared near significant market turning points, particularly during oversold conditions following the market crash in late 2008.

3.1.4 Evaluating the Performance of MACD and RSI Indicators on Brent Crude Oil Index During the Global Financial Crisis

During the financial crisis, crude oil prices have fluctuated significantly. Brent crude oil reached historic highs in mid-2008 before declining rapidly as global economic activity slowed and expectations for demand has decreased.

From 2008 to 2009, the MACD indicator generated 21 indications. While several of these indications identified short-term market trends, the extraordinary volatility in oil prices referred to several signals were immediately invalidated by the price reversals.

In comparison to the RSI indicator that on its behalf have produced four signals in the same time duration. Although fewer in number of signals but these signals emerged during periods of a high volatility in the prices and were frequently connected with major turning moments in the oil market.

3.1.5 Evaluating the Performance of MACD and RSI Indicators on Nasdaq-100 Index Index During the Global Financial Crisis

During the financial crisis, the Nasdaq-100 index tracked the broader equities market, but with substantially higher volatility due to its heavy concentration in technology-related companies. During this time, the MACD indicator produced 18 signals. While some of these signals correctly anticipated short-term trend reversals, others demonstrated the difficulties of using momentum-based indicators in circumstances typified by intense and rapidly changing market conditions.

On the other hand, the RSI indicator generated only four alerts throughout the same time period. Consistent with observations across other asset classes, RSI signals were less frequent because the indicator requires prices to reach more extreme momentum levels before generating overbought or oversold conditions. During times of increased market volatility, RSI tended to provide fewer but more selective signals.

3.1.6 Summary of Indicator Performance

Table 1 Summarizes the performance of both indicators across the analyzed assets during the Global Financial Crisis.

Asset	Indicator	Total signals	Correct signals	Accuracy
Gold	MACD	20	16	80.0%
Gold	RSI	7	5	71.0%
S&P 500	MACD	23	16	69.6%
S&P 500	RSI	5	3	60.0%
Brent Crude Oil	MACD	21	11	52.4%
Brent Crude Oil	RSI	4	3	75.0%
Nasdaq-100	MACD	18	10	55.6%
Nasdaq-100	RSI	4	2	50.0%

According to the data collected, the MACD indicator have provided a significantly greater number of trading signals across all investigated assets than the RSI indicator. The result is consistent with MACD's structural architecture, which responds more rapidly to changes in price momentum and the evolution of market trends.

In comparison, the RSI indicator provided a fewer alerts because it requires market movements to reach specified overbought or oversold levels before initiating potential trade signals. As a result, while RSI indications were less common, they tended to occur near major market turning moments, where momentum fatigue or potential reversals were more likely.

According to the data, the MACD indicator provided a significantly greater number of trading signals across all investigated assets than the RSI indicator. This conclusion is consistent with MACD's structural architecture, which responds more rapidly to changes in price momentum and the evolution of market trends.

On the other hand and in comparison, the RSI indicator produced fewer alerts because it requires market movements to reach preset overbought or oversold levels before triggering potential trade signals. As a result, while RSI indications were less common, they tended to occur near major market turning moments, where momentum fatigue or potential reversals were more likely.

3.1.7 Indicator Accuracy Comparison During the Global Financial Crisis

Figure 1 Accuracy of MACD and RSI signals across assets during the Global Financial Crisis

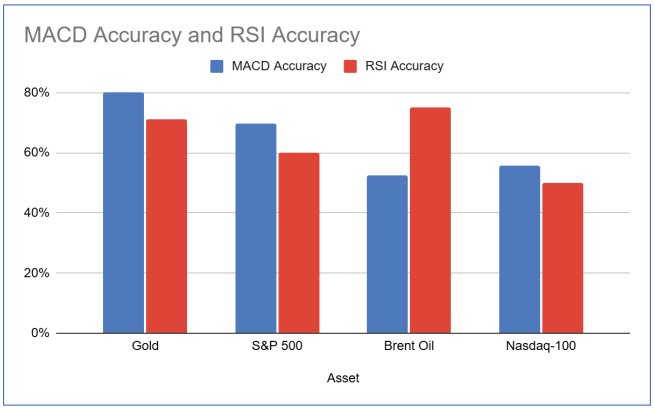


Figure 1 above shows the accuracy comparison of signals for both indicators the MACD and RSI signals across the four assets evaluated during the Global Financial Crisis. The results reveal that the MACD indicator created more signals, whereas the RSI produced fewer but more accurate indications, especially in commodity markets.

3.1.8 Signal Frequency Comparison During the Global Financial Crisis

Table 2 Number of trading signals generated by MACD and RSI during the Global Financial Crisis

Asset	MACD signals	RSI signals
Gold	20	7
S&P 500	23	5
Brent Oil	21	4
Nasdaq-100	18	4

RSI throughout the global financial crisis. MACD provided have idicated alot more signals and as a result of its sensitivity to the momentum changes, whereas on the other hand the indicator RSI provided less signals given that it requires extreme overbought or oversold conditions beofre stimulating the trades.

3.2 COVID-19 Market Crisis (2020)

The COVID-19 pandemic caused one of the most violent market crashes throughout the modern financial history in fact, the pandemic have started in February 2020 that's when the global financial markets became more volatile specially when the governments around the world have implemented a lockdown measures and economic activity slowed significantly therefore the equity markets have recorded a significant decline in March 2020, in contrast the commodities and safe-haven investments responded severely to the increased uncertainty that the global health crisis have caused.

However; the same methodology as described in the previous section was used to study and evaluate the behavior of technical indicators throughout this time period. Trading signals created by the Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) indicators were identified using daily price data from January to September 2020.

Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) indicators were identified on daily price charts between January 2020 and September 2020 for four major assets: gold, the S&P 500 index, the Nasdaq-100 index, and Brent crude oil.

Each signal was evaluated using the ten-day price comparison method described earlier. A bullish signal was considered successful if the price increased after ten trading days, while a bearish signal was considered successful if the price declined over the same period.

3.2.1 Evaluating the Performance of MACD and RSI Indicators on Gold During the COVID-19 Market Crisis

During the COVID-19 pandemic crisis, investors demanded protection from financial market uncertainty and that's where the gold played a crucial part. During the period that is under examination, the MACD indicator generated five trade signals. Several of these signals correctly predicted short-term directional fluctuations in gold prices as the market responded to shifts in investor mood and monetary policy expectations.

Also, throughout the same timeframe, the RSI indicator sent out three signals. As seen in the last crisis analysis, RSI produced less signals because it relies on extreme overbought or oversold situations to provide a trading signal. Despite the modest number of signals, numerous RSI indications were associated with significant price corrections in the gold market.

3.2.2 Evaluating the Performance of MACD and RSI Indicators on S&P 500 Index During the COVID-19 Market Crisis

During the initial COVID-19 pandemic crisis, the S&P 500 recorded one of its fastest declines in history, followed by a rapid recovery supported by large-scale monetary and fiscal stimulus.

During this point in time, the MACD indicator generated four trading signals. Several of these signals have reflected the significant directional shifts that occurred during the March 2020 market meltdown and following recovery phase, whereas when comparing that to the RSI indicator generated four alerts. Because of the quick rate of market moves during the crisis, several RSI indications arrived after large price changes had already occurred. This highlights the difficulties associated with using oscillators in extremely volatile market conditions.

3.2.3 Evaluating the Performance of MACD and RSI Indicators on Brent Crude Oil During the COVID-19 Market Crisis

The crude oil market faced extreme volatility during the COVID 19 epidemic back when the global demand has decreased in response to all the travel restrictions that have occured and the worldwide economic shutdowns, the Oil prices decreased significantly in early 2020 before stabilizing later in that year.

During that particular time the MACD indicator has generated

six indications. although many of these signals has accurately predicted short-term market movements, the high volatility in oil prices often resulted in an unexpected reversal on its behalf reducing the predictive reliability of some signals. whereas the RSI indicator produced four signals throughout the same time period and these indications came around extreme market conditions and were linked to important price turning moments following the sharp drop in oil prices in the first half of 2020.

3.2.4 Evaluating the Performance of MACD and RSI Indicators on Nasdaq - 100 During the COVID-19 Market Crisis

The Nasdaq-100 index noticed significant volatility in the early stages of the COVID-19 crisis, but recovered quickly as technology businesses benefited from increased digital activity throughout the epidemic.

so first; The MACD indicator generated ten trade signals between January and September 2020 yet Several of these indications reflected the technology sector's strong upward momentum following the initial market collapse, on the other hand we study the RSI indications in which RSI indicator produced seven signals throughout the same time period. As with other assets, RSI signals were less frequent and more likely to emerge amid extreme market situations.

3.2.5 Summary of Indicator Performance

Table 3 The performance of both indicators during the COVID-19 crisis

Asset	Indicator	Total signals	Correct signals	Accuracy
Gold	MACD	5	3	60.0%
Gold	RSI	3	2	66.7%
S&P 500	MACD	4	3	75.0%
S&P 500	RSI	4	2	50.0%
Brent Crude Oil	MACD	6	4	66.7%
Brent Crude Oil	RSI	4	2	50.0%
Nasdaq-100	MACD	10	8	80.0%
Nasdaq-100	RSI	7	3	42.9%

Author's calculations based on TradingView data (2026)

The outcomes demonstrate that the MACD indicator generated more trading signals across most assets than the RSI and this indicates the MACD's sensitivity to momentum shifts and short-term trend creation.

In comparison by the RSI indicator which generated less signals because it requires the market to reach extreme momentum conditions before giving trade signals and when the COVID-19 crisis happened the rapidity of market moves reduced the effectiveness of RSI signals, especially in equities markets where abrupt reversals happened in short time intervals. Therefore, the results indicate that momentum-based indicators like MACD may respond more effectively to rapid market changes, whilst oscillators like RSI may perform better in clearer mean-reversion conditions.

3.2.6 Accuracy of MACD and RSI Trading Signals During the COVID-19 Market Crisis

Figure 2 Indicator accuracy comparison signals during the COVID-19 market crisis

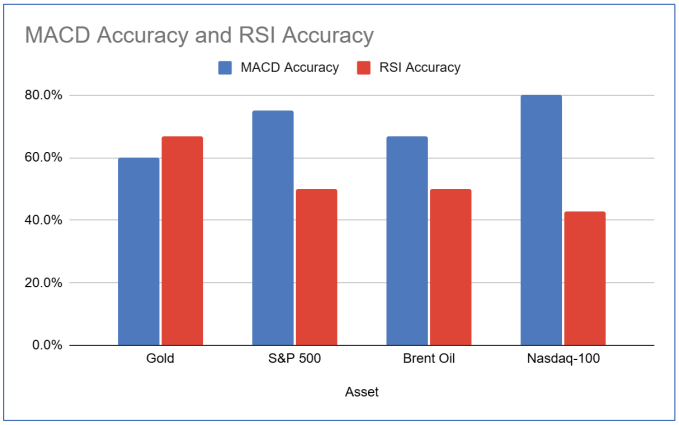


Figure 2 compares the accuracy of MACD and RSI trading signals across the assets studied during the COVID-19 market crisis. The results demonstrate that MACD performed better in equities markets during this period of high volatility, apart from RSI which has produced fewer signals and had lower predictive accuracy across multiple assets.

3.2.7 Signal Frequency Comparison (COVID)

Table 4 Number of MACD and RSI signals generated during the COVID-19 Crisis

Asset	MACD signals	RSI signals
Gold	5	3
S&P 500	4	4
Brent Oil	6	4
Nasdaq-100	10	7

Author's calculations based on TradingView data (2026)

Table 4 represents the total number of trade signals produced by the MACD and RSI indicators during the COVID-19 crisis. MACD provided more indications due to its sensitivity to momentum changes, although RSI generated less signals since it requires extreme overbought or oversold levels to initiate trades.

3.3 Inflation Shock and Geopolitical Tensions (2022)

The year 2022 was characterized by a unique combination

of macroeconomic and geopolitical shocks that significantly affected global financial markets. Inflation rates in major economies reached multi decade highs, prompting central banks to implement aggressive monetary tightening policies. At the same time, the outbreak of the Russia–Ukraine conflict in February 2022 triggered major disruptions in global energy and commodity markets. These developments generated heightened volatility across equity, commodity, and currency markets, making 2022 an important period for evaluating the performance of technical indicators under conditions of elevated macroeconomic uncertainty.

3.3.1 Summary of Indicator Performance During the Inflation Shock and Geopolitical Tensions (2022)

Table 5 Indicator performance during the inflation shock and geopolitical tensions (2022)

Asset	Indicator	Total signals	Correct signals	Accuracy
Gold	MACD	8	4	50.00%
Gold	RSI	3	2	66.70%
S&P 500	MACD	9	5	55.60%
S&P 500	RSI	3	2	66.70%
Brent Crude Oil	MACD	9	4	44.40%
Brent Crude Oil	RSI	2	2	100.00%
Nasdaq-100	MACD	12	5	41.70%
Nasdaq-100	RSI	3	1	33.30%

Author's calculations based on TradingView data (2026)

The results presented in Table 5 summarize the performance of the MACD and RSI indicators across the analyzed assets during the Inflation Shock and Geopolitical Tensions period in 2022. The MACD indicator generated a higher number of trading signals compared with RSI across all assets, reflecting its sensitivity to changes in price momentum and trend development.

In contrast, the RSI indicator produced fewer signals because it requires the market to reach extreme overbought or oversold conditions before generating trading signals. As a result, RSI signals occurred less frequently but occasionally aligned with significant market turning points.

Indicator performance during the 2022 crisis was generally lower compared with earlier crisis periods analyzed in this study. The persistent macroeconomic trends driven by aggressive monetary tightening, elevated inflation, and geopolitical uncertainty created prolonged directional market movements, which reduced the reliability of short-term momentum signals.

Commodity markets such as gold and Brent crude oil demonstrated relatively stronger RSI performance, while equity indices including the S&P 500 and Nasdaq-100 exhibited lower predictive accuracy. This reflects the difficulty of applying technical momentum indicators during sustained macro-driven market adjustments.

3.3.2 Indicator Accuracy Comparison Across Assets (2022 Crisis)

Figure 3 Accuracy of MACD and RSI signals during the 2022 inflation shock and geopolitical crisis period

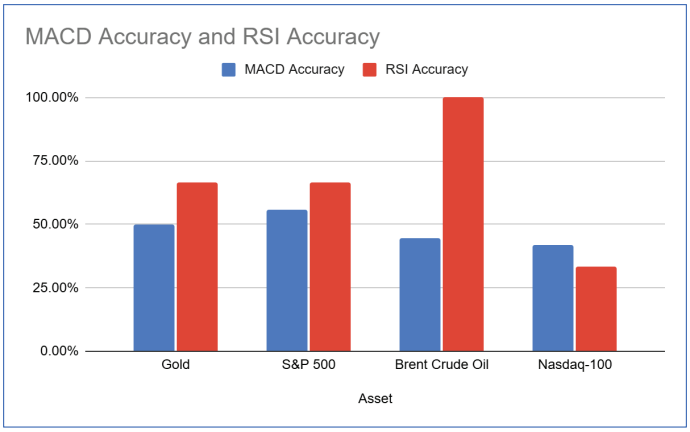


Figure 3 compares the predictive accuracy of MACD and RSI signals during the 2022 inflation shock and geopolitical crisis period. While MACD generated a larger number of signals, RSI occasionally demonstrated higher accuracy due to its ability to capture extreme market conditions. However, the limited number of RSI signals reduces its responsiveness to frequent market movements.

3.3.3 Signal Frequency Comparison During the Inflation Shock (2022)

Table 6 Indicator performance during the inflation shock and geopolitical tensions (2022)

Asset	MACD signals	RSI signals
Gold	8	3
S&P 500	9	3
Brent Crude Oil	9	2
Nasdaq-100	12	3

Author's calculations based on TradingView data (2026)

Table 6 illustrates the number of trading signals generated by the MACD and RSI indicators across the analyzed assets during the 2022 inflation and geopolitical shock period. As expected, the MACD indicator produced a significantly higher number

of signals compared with RSI due to its greater sensitivity to changes in price momentum.

3.4 Banking Sector Stress and Monetary Policy Uncertainty (2023)

back in 2023 the financial markets faced increasing volatility as a consequence of the stress that has occurred in the global banking and the continuous monetary policy uncertainty and during the collapse of Silicon Valley Bank in March 2023 and alongside the subsequent instability across many financial institutions, sparked severe reactions in the equities, commodities, and safe-haven markets. At the same time, central banks maintained restrictive monetary conditions in response to persistent inflation, increasing the investor's uncertainty and these developments resulted in a period of increased market volatility, making it important to assess the performance of technical indicators in the context of renewed financial stress.

3.4.1 Summary of Indicator Performance During Banking Sector Stress and Monetary Policy Uncertainty (2023)

Table 7 Summary of indicator performance during the banking sector stress (2023)

Asset	Indicator	Total signals	Correct signals	Accuracy
Gold	MACD	7	3	42.90%
Gold	RSI	2	1	50.00%
S&P 500	MACD	6	4	66.70%
S&P 500	RSI	2	1	50.00%
Brent Crude Oil	MACD	7	2	28.60%
Brent Crude Oil	RSI	3	2	66.70%
Nasdaq-100	MACD	9	4	44.40%
Nasdaq-100	RSI	3	1	33.30%

Author's calculations based on TradingView data (2026)

As a start, involving with the previous crises evaluated in this study, the MACD indicator produced more signals than the RSI where on its behalf highlights the momentum-based aspect of the MACD indicator that reacts quickly to shifts in trend direction, also during the banking stress phase the markets experienced regular momentum shifts as investor sentiment changed frequently in response to developments in the financial sector.

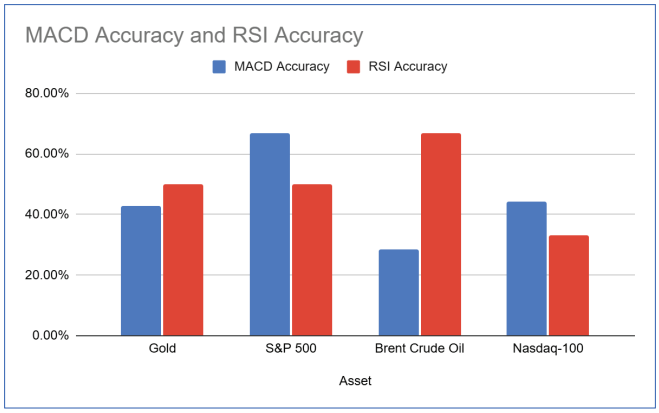
Also, RSI signals remained restricted across all assets and this is because the RSI indicator required prices to reach extreme levels before triggering signals, and it generated less possibilities during a time that is distinguished by trend modifications than during repeated overbought or oversold circumstances.

Moving forward, performance changed slightly between asset classes. Gold, which is often known as a safe-haven asset during times of financial instability have demonstrated relatively low indication accuracy as prices reacted strongly to banking sector developments. Equity indices that included the S&P 500 and Nasdaq-100 were showing a mixed performance showing the impact of both banking sector concerns and evolving expectations for central bank policies.

Nevertheless; Brent crude oil had the lowest MACD accuracy among the assets that has been studied in this research paper. During the banking sector crisis the oil prices were not just influenced by the financial market sentiments, but also by supply-side variables and global economic expectations, reducing the efficiency of momentum based indicators.

3.4.2 Indicator Accuracy Comparison Assets During the 2023 Banking Stress

Figure 4 Indicator accuracy comparison across assets auring the 2023 banking stress



The forecasting accuracy of the MACD and RSI indicators across the examined assets all through the 2023 banking sector stress period is shown in Figure 4.8 above. where the comparison reveals that different asset, classes had different levels of indicator efficacy, which reflects variations in the market's structure and the type of price fluctuations that occurred during this phase.

In comparison to the other assets under the study and analysis i this research paper the researcher addressed that the S&P 500 showed relatively stronger MACD performance, indicating that momentum-based signals were more successful in identifying directional changes in the overall equities market during this time and on the other hand the gold has performed moderately on both measures, which is in line with its function as a safe-haven asset responding to changes in investor sentiment and financial instability.

The MACD performance of Brent crude oil was weaker during this time, which indicated that that supply-side changes and broader macroeconomic expectations had an impact on oil prices in addition to short-term market momentum. RSI recorded more extreme market turning times, therefore even if its signals were less frequent, it demonstrated somewhat more reliable performance in some assets.

Therefore, when taking all together the data indicates that, during the banking stress period, RSI offered fewer but occasionally more selective signals, whereas MACD produced more opportunities and its predictive success was highly dependent on the asset under analysis.

3.4.3 Signal Frequency Comparison During the 2023 Banking Sector

Table 8 Signal frequency comparison of MACD and RSI during the 2023 banking sector stress

Asset	MACD signals	RSI signals
Gold	7	2
S&P 500	6	2
Brent Crude Oil	7	3
Nasdaq-100	9	3

Author's calculations based on TradingView data (2026)

Table 8 compares the number of trading signals generated by the MACD and RSI indicators across the analyzed assets during the 2023 banking sector stress period. The results demonstrate that the MACD indicator delivered significantly more signals than the RSI in all asset classes which reveals how the MACD's reflect higher sensitivity to short-term changes in price momentum and trend direction. however, the higher signal frequency observed in MACD shows that markets experienced repeated momentum volatility during this time as investors reacted to banking sector developments and changing in expectations regarding the monetary policy, compared to the RSI indicator which have delivered fewer signals and this can be explained because it requires prices to reach more extreme overbought or oversold levels before a signal is issued.

in short, the data shows that MACD was more responsive to the variations that defined the 2023 banking sector stress environment, but RSI was more selective, producing indications only during stronger market extremes.

3.5 Cross-Crisis Indicator Performance Comparison

Table 9 Cross crisis indicator accuracy across market crises

Crisis	MACD Average Accuracy	RSI Average Accuracy
Global Financial Crisis	64.4%	64.0%
COVID-19 Shock	56.3%	62.5%
Inflation & Geopolitical Shock	47.9%	66.7%
Banking Sector Stress	45.7%	50.0%

Author's calculations based on TradingView data (2026)

The cross-crisis analysis reveals significant changes in the MACD and RSI indicators' performance across market circumstances. throughout the Global Financial Crisis, both indicators showed relatively high prediction accuracy, indicating the presence of long-term market trends and rapid price reversals.

Despite this, during the COVID-19 market shock, RSI had slightly better predictive performance because the intense market fluctuations caused more frequent overbought and oversold positions. In contrast, the Inflation and Geopolitical Shock era of 2022 showed reduced MACD accuracy, as markets were highly influenced by macroeconomic events and policy uncertainty rather by consistent momentum patterns.

The 2023 banking sector stress incident had the lowest MACD accuracy of the crises examined in this research paper, demonstrating the complex and fast changing nature of market movements during this time period. Perhaps the comparison indicates that indicator effectiveness varies with market volatility and the underlying determinants of price fluctuations.

DISCUSSION

4.1 Overview of Findings

This research paper examined the predictive performance of two widely used technical indicators, Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) in across multiple financial assets during four periods of market crisis: The Global Financial Crisis (2008-2009), the COVID-19 market shock (2020), the inflation and geopolitical volatility of 2022, and the banking sector crisis occurring back in 2023.

the key findings from the analyzing the result of this study demonstrate that the effectiveness of technical indicators differs based on the market situation, asset class, and the underlying sources of volatility. but certainly, the MACD indicator generated more trade signals during all crises, demonstrating its sensitivity to changes in price momentum and trend direction. The results of this research paper can be compared to previous research by (Melda and Mira 2025) in which the findings conclude that the Relative Strength Index (RSI) is more ideal for identifying overbought and oversold market conditions, whereas the Moving Average Convergence Divergence (MACD) is effective in detecting market momentum and longer-term trend direction whereas the results of this study are partially consistent with these findings the RSI indications demonstrated less frequently during the crisis periods investigated since the indicator required severe momentum levels to produce overbought or oversold conditions. on the other hand, the MACD indicator have generated more signals since it reacts faster to shifts in market momentum and trend direction.

However; Across the major crises that have occurred during the past two decades, MACD frequently responded faster to swings in market momentum, although RSI tended to capture more selective turning points during moments of dramatic price volatility.

4.2 Indicator Behavior Across Different Market Conditions

The analysis reveals that market structure and the nature of volatility have a significant impact on indicator performance, especially during the Global Financial Crisis, markets had experienced strong directional trends and quick reversals, making MACD and RSI reasonably good predictors. During this time, momentum-based indicators were able to record considerable price swings due to the presence of strong market cycles.

During the COVID-19 pandemic market shock and the market swings caused quick shifts between oversold and overbought levels. In such cases, RSI signals became more effective at detecting key turning moments, although MACD continued to give numerous alerts in reaction to big momentum fluctuations.

Compared to that, inflation-driven volatility in 2022 created a more complicated trading environment, with markets influenced not only by price momentum but also by larger macroeconomic issues like as monetary policy tightening, rising interest rates, and geopolitical developments. These structural factors impacted the effectiveness of only using momentum-based indicators, resulting in lower forecast accuracy for MACD across several assets. A similar pattern was demonstrated during the 2023 banking sector stress incident, when financial markets were mostly driven by fluctuations in investor mood and changing assumptions about the financial system's viability. The quick and sentiment-driven character of these moves resulted in short-term price changes, reducing the reliability of momentum signals across many asset classes.

4.3 Differences Across Asset Classes

The research results also demonstrate substantial differences in indicator performance across asset types. Gold, which recognized as safe-haven asset during times of financial stress, performed in balance between the two indicators. During market stress, gold prices can react significantly to fluctuations in risk perception and increased investor demand for defensive assets. These circumstances generate clearer and more lasting technical patterns, allowing momentum indicators to capture price signals more consistently than in more turbulent market environments.

On the other hand, stock indices such as the s&p 500 and Nasdaq-100 had more volatile indicator performance. Macroeconomic expectations, corporate earnings, and monetary policy changes all have an impact on stock markets. These factors can cause complex price dynamics, which reduces the effectiveness of basic momentum-based signals. However; Brent crude produced the most inconsistent results among the assets studied therefore its shown that the oil prices are significantly influenced by supply-demand imbalances, geopolitical upheavals, and fluctuations in global economic activity, resulting in abrupt and erratic price movements that are difficult to detect using typical technical indicators.

4.4 Implications for Traders and Market Participants

The findings of this research show that traders and analysts should take consideration of the larger market situation while using technical indicators. While MACD's responsiveness to price momentum may result in more frequent trading opportunities, the accuracy of these signals differs depending on the asset class and market conditions.

Although the Relative Strength Index (RSI) indicator delivers signals less frequently, it may be especially effective in identifying significant turning points during periods of extreme market conditions. Integrating RSI with other indicators, such as MACD, or combining it with additional analytical tools, might improve trading methods, especially during times of increased market uncertainty. Thus, the results we obtained demonstrate that technical indicators should not be used in isolation; rather, their usefulness improves when understood within the larger economic and market context, allowing for more informed decision-making in volatile financial situations.

4.5 Limitations of the Study

Despite providing useful insights concerning indicator performance during various crises, this study has some limitations. The analysis focuses on a small number of assets and variables that has been selected, which may not adequately represent the diverse variety of instruments available in global financial markets.

Future study could overcome these constraints by broadening the analysis to include and study more technical indicators, longer time periods, and a wider range of asset classes.

CONCLUSION

5.1 Summary

The aim of this research paper is to test the accuracy and effectiveness of two widely used technical indicators, the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI), across four major periods of financial market stress: the Global Financial Crisis (2008–2009), the COVID-19 market shock (2020), the inflation and geopolitical volatility of 2022, and the banking sector stress of 2023. Using daily price data from gold, the S&P 500 index, the Nasdaq-100 index, and Brent crude oil, the research assessed trading signals that are generated by both indicators through a ten-day forward price comparison method.

to conclude all above; it was founded that the performance of technical indicators is heavily influenced by both the dynamics of the market environment and the structure of the issue that lies beneath it, in other words, during the Global Financial Crisis, both MACD and RSI demonstrated strong predictive performance, indicating the presence of distinct directional trends and sharp market reversal. In contrast, the quick and unusual market fluctuations experienced during the COVID-19 shock made technical signals more difficult to interpret, however momentum-based indicators such as MACD were nevertheless able to detect many key shifts in market direction.

During the banking sector crisis back in the year 2023 and the inflation-driven volatility of 2022, which were identified by significant macroeconomics factors, monetary policy uncertainty, and sentiment-driven market behavior, indicator reliability decreased. In these conditions, the dominant role of macroeconomic forces reduced the predicted accuracy of

technical signals and diminished the efficacy of short-term momentum indicators.

Due to its sensitivity to shifts in price momentum, MACD produced more trade signals across all crisis periods, but RSI produced fewer but more selective signals that occasionally coincided with important market turning points during times of extreme volatility. The research paper have also emphasizes how its common that asset type affect indicator performance, for example the safe haven which is the gold demonstrated relatively more stable indicator behavior compared with equity indices and crude oil, reflecting its function as what its known to be viewed as the safe-haven asset during periods of financial uncertainty.

In short, the findings of this research paper indicate that technical indicators respond differently depending on the sort of market shock and the structural dynamics influencing price fluctuations, however while continuous macroeconomic trends may reduce the efficacy of momentum based indicators, crisis scenarios marked by swift reversals and high volatility may provide more favorable settings. These results highlight how crucial it is to combine technical indicators with more comprehensive macroeconomic and market background when examining financial markets during stressful times.

5.2 Practical Implication

The findings of this research highlight the importance of understanding the market environment when applying technical indicators. While MACD may provide more frequent signals and respond quickly to shifts in price momentum, its predictive accuracy may decline during periods dominated by macroeconomic policy changes or structural market trends.

RSI, although less active, may offer value in identifying extreme market conditions and potential turning points. Traders and analysts may therefore benefit from combining multiple indicators rather than relying on a single technical signal.

5.3 Future Research

The upcoming researches in the future may aim to expand the scope of the analysis by integrating and studying more indicators, time periods, or different evaluation perspectives. Combining technical analysis with macroeconomic variables or machine learning technologies may provide further insights into how trading signals function in various market scenarios.

5.4 Final Remarks

All in all, this research paper demonstrates that technical indicators such as MACD and RSI can be quite useful insights during time of financial market crisis, but their effectiveness is not constant across all environments. Market structure, volatility dynamics, and the nature of the overall crisis all influence how technical signals appear.

By studying multiple crises and instruments classes, this research paper contributes to a better understanding of the performance of the traditional technical indicators under extreme market conditions. These findings might help traders, analysts, and investors apply technical analysis more effectively by considering both indicator signals and the broader economic context in which occur.

REFERENCES

Cheema MA, Faff R, Szulczyk K. 2022. The 2008 global financial crisis and COVID-19 pandemic: How safe are the safe haven assets? International Review of Financial Analysis. 83.

Investopedia. 2024. Technical analysis. Available from: https://www.investopedia.com/terms/t/technicalanalysis.asp

Melda M, Mira S. 2025. Analysis of the effectiveness of RSI and MACD indicators in addressing stock price volatility. Wiga: Journal of Research and Business Economics. Available from: https://www.researchgate.net/publication/392317792_Analysis_of_the_Effectiveness_of_RSI_and_MACD_Indicators_in_Addressing_Stock_Price_Volatility

Moodley F. 2026. The impact of market dynamics and geopolitical uncertainty on BRICS property market returns. Risks. 14(3). Available from: https://www.mdpi.com/2227-9091/14/3/55

Sari IY, Perdana AA, Zuhra S. 2025. A comparative study of MACD and RSI indicators in identifying energy sector stock trends. Journal of Economic Research. 6(4). Available from: https://www.iicls.org/index.php/jer/article/view/1559

Singh K, Priyanka. 2025. Unlocking trading insights: A comprehensive analysis of RSI and MA indicators. Metamorphosis: A Journal of Management Research. 24(2):164–172. Available from: https://journals.sagepub.com/doi/10.1177/09726225241310978

TradingView. 2026. Financial market charts and historical price data. Available from: https://www.tradingview.com

Zakry M, Ben Hssain L, Agouram J, Lakhnati G. 2026. Dynamic linkages between global commodity prices and stock markets in times of crisis: Evidence from a DCC-GARCH framework. Scientific African. 31. Available from: https://www.sciencedirect.com/science/article/pii/S2468227626000451

APPENDIX A

Trading Signals Dataset Used in Indicator Evaluation

This chapter presents the dataset of MACD and RSI trading signals used in the empirical analysis. Historical price data for the analyzed assets were obtained from TradingView aming the years 2008 to 2023 (TradingView 2026), and the signals were identified and recorded by the author.

Table A1. MACD trading signals for gold during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Gold	Global Financial Crisis	MACD	Bear-ish	31-1-2008	905.00	943.45
Gold	Global Financial Crisis	MACD	Bullish	19-2-2008	943.45	972.80
Gold	Global Financial Crisis	MACD	Bear-ish	17-3-2008	982.30	902.80
Gold	Global Financial Crisis	MACD	Bear-ish	17-4-2008	918.00	868.17
Gold	Global Financial Crisis	MACD	Bullish	13-5-2008	864.06	886.70
Gold	Global Financial Crisis	MACD	Bear-ish	17-6-2008	954.80	919.50
Gold	Global Financial Crisis	MACD	Bear-ish	21-7-2008	945.85	872.80
Gold	Global Financial Crisis	MACD	Bullish	15-9-2008	779.37	836.53
Gold	Global Financial Crisis	MACD	Bear-ish	13-10-2008	836.69	738.68
Gold	Global Financial Crisis	MACD	Bullish	12-11-2008	743.75	800.15
Gold	Global Financial Crisis	MACD	Bullish	6-1-2009	842.63	894.68
Gold	Global Financial Crisis	MACD	Bullish	22-1-2009	894.68	939.03
Gold	Global Financial Crisis	MACD	Bear-ish	25-2-2009	952.51	915.08
Gold	Global Financial Crisis	MACD	Bullish	22-4-2009	903.08	922.36
Gold	Global Financial Crisis	MACD	Bear-ish	8-6-2009	945.45	939.20
Gold	Global Financial Crisis	MACD	Bullish	15-7-2009	936.90	967.40
Gold	Global Financial Crisis	MACD	Bullish	31-8-2009	956.90	1012.10
Gold	Global Financial Crisis	MACD	Bullish	5-10-2009	1041.20	1051.40
Gold	Global Financial Crisis	MACD	Bullish	2-11-2009	1083.80	1144.50
Gold	Global Financial Crisis	MACD	Bear-ish	3-12-2009	1160.00	1087.60

Table A2. RSI trading signals for gold during the Global Financial Crisis (2008–2009)

As-set	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Gold	Global Financial Cri-sis	RSI	Bear-ish	14-1-2008	891.90	925.90
Gold	Global Financial Cri-sis	RSI	Bear-ish	30-1-2008	924.10	927.40
Gold	Global Financial Cri-sis	RSI	Bear-ish	13-3-2008	1002.60	904.10
Gold	Global Financial Cri-sis	RSI	Bullish	14-8-2008	787.30	801.29
Gold	Global Financial Cri-sis	RSI	Bullish	10-9-2008	746.80	869.05
Gold	Global Financial Cri-sis	RSI	Bear-ish	19-2-2009	993.63	907.03
Gold	Global Financial Cri-sis	RSI	Bear-ish	1-6-2009	981.63	933.33
Gold	Global Financial Cri-sis	RSI	Bear-ish	2-12-2009	1207.60	1083.60

Table A3. MACD trading signals for the S&P 500 during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indicator	Bearish Signal	Date	Price	Price after 10 days
S&P 500	Global Financial Crisis	MACD	Bullish	29-1-2008	1362.30	1348.90
S&P 500	Global Financial Crisis	MACD	Bearish	4-3-2008	1326.80	1330.70
S&P 500	Global Financial Crisis	MACD	Bullish	19-3-2008	1298.40	1369.30
S&P 500	Global Financial Crisis	MACD	Bullish	16-4-2008	1364.70	1385.60
S&P 500	Global Financial Crisis	MACD	Bearish	20-5-2008	1413.40	1337.20
S&P 500	Global Financial Crisis	MACD	Bullish	17-7-2008	1260.30	1267.20
S&P 500	Global Financial Crisis	MACD	Bearish	2-9-2008	1277.60	1213.60
S&P 500	Global Financial Crisis	MACD	Bearish	24-9-2008	1050.90	1029.50
S&P 500	Global Financial Crisis	MACD	Bearish	28-10-2008	940.50	899.00
S&P 500	Global Financial Crisis	MACD	Bullish	25-11-2008	857.40	899.20

S&P 500	Global Financial Crisis	MACD	Bearish	12-1-2009	870.30	845.70
S&P 500	Global Financial Crisis	MACD	Bearish	13-2-2009	826.80	700.80
S&P 500	Global Financial Crisis	MACD	Bullish	11-3-2009	721.40	813.90
S&P 501	Global Financial Crisis	MACD	Bearish	12-5-2009	908.40	893.10
S&P 500	Global Financial Crisis	MACD	Bearish	12-6-2009	946.20	918.90
S&P 500	Global Financial Crisis	MACD	Bullish	14-7-2009	905.80	979.60
S&P 500	Global Financial Crisis	MACD	Bearish	12-8-2009	1005.80	1028.10
S&P 500	Global Financial Crisis	MACD	Bearish	28-8-2009	1028.90	1028.60
S&P 500	Global Financial Crisis	MACD	Bearish	24-9-2009	1050.80	1065.90
S&P 500	Global Financial Crisis	MACD	Bullish	9-10-2009	1071.50	1079.90
S&P 500	Global Financial Crisis	MACD	Bearish	23-10-2009	1075.50	1069.30
S&P 500	Global Financial Crisis	MACD	Bearish	27-11-2009	1091.50	1106.40
S&P 500	Global Financial Crisis	MACD	Bullish	23-12-2009	1120.00	1145.00

Table A4. RSI trading signals for the S&P 500 during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
S&P 500	Global Financial Crisis	RSI	Bullish	16-7-2008	1245.4	1284.3
S&P 500	Global Financial Crisis	RSI	Bear-ish	10-10-2008	899.2	878.8
S&P 500	Global Financial Crisis	RSI	Bullish	9-3-2009	676.5	822.9
S&P 500	Global Financial Crisis	RSI	Bear-ish	10-8-2009	1007.1	1025.6
S&P 500	Global Financial Crisis	RSI	Bear-ish	22-9-2009	1071.7	1054.7

Table A5. MACD trading signals for Brent Crude Oil during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indica-tor	Bearish Signal	Date	Price	Price after 10 days
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	9-1-2008	94.37	91.83
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	8-2-2008	91.94	99.47
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	19-3-2008	100.72	102.59
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	7-4-2008	107.14	116.46
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	5-5-2008	117.99	132.7
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	28-5-2008	130.93	135.02
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	7-7-2008	136.43	125.29
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	21-8-2008	120.16	106.3
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	2-9-2008	108.34	95.19
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	19-9-2008	99.61	90.25
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	2-10-2008	90.56	72.03
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	31-10-2008	65.32	51.84
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	13-2-2009	44.81	46.35

Brent Crude Oil	Global Financial Crisis	MACD	Bearish	17-4-2009	53.35	52.85
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	18-6-2009	71.06	66.65
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	21-7-2009	66.87	71.7
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	18-8-2009	72.37	76.73
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	17-9-2008	71.55	69.19
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	2-10-2009	68.07	77.24
Brent Crude Oil	Global Financial Crisis	MACD	Bearish	2-11-2009	76.55	79.47
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	23-12-2009	75.45	81.73

Table A6. RSI trading signals for Brent Crude Oil during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Brent Crude Oil	Global Financial Crisis	RSI	Bearish	26-5-2008	132.37	133.91
Brent Crude Oil	Global Financial Crisis	RSI	Bullish	17-9-2008	94.84	95.33
Brent Crude Oil	Global Financial Crisis	RSI	Bullish	5-12-2008	39.74	41.45
Brent Crude Oil	Global Financial Crisis	RSI	Bearish	16-6-2009	71.06	66.65

Table A7. MACD trading signals for the Nasdaq-100 Index during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indica-tor	Bearish Signal	Date	Price	Price after 10 days
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	1-1-2008	2040.40	1872.30
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	31-1-2008	1841.40	1787.60
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	12-3-2008	1731.10	1777.90
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	14-4-2008	1790.90	1923.30
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	20-5-2008	2000.70	2021.50
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	16-7-2008	1843.90	1852.80
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	26-7-2008	1886.30	1738.10
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	27-10	1169.80	1251.00
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	25-11-2008	1142.60	1222.40
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	13-1-2009	1202.70	1235.90

Nasdaq - 100	Global Financial Crisis	MACD	Bearish	13-2-2009	1239.90	1074.70
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	11-3-2009	1126.00	1236.90
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	7-5-2009	1389.80	1367.40
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	12-6-2009	1490.00	1480.20
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	7-8-2009	1619.50	1637.80
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	9-9-2009	1669.20	1637.80
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	14-10-2009	1754.30	1685.10
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	9-11-2009	1768.40	1792.90
Nasdaq - 100	Global Financial Crisis	MACD	Bearish	27-11-2009	1765.50	1792.10
Nasdaq - 100	Global Financial Crisis	MACD	Bullish	21-12-2009	1828.80	1878.40
Brent Crude Oil	Global Financial Crisis	MACD	Bullish	23-12-2009	75.45	81.73

Table A8. RSI trading signals for the Nasdaq-100 Index during the Global Financial Crisis (2008–2009)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Nasdaq - 100	Global Financial Crisis	RSI	Bullish	23-1-2008	1789.50	1739.00
Nasdaq - 100	Global Financial Crisis	RSI	Bullish	10-10-2008	1269.80	1202.30
Nasdaq - 100	Global Financial Crisis	RSI	Bearish	5-8-2009	1614.10	1569.60
Nasdaq - 100	Global Financial Crisis	RSI	Bearish	23-9-2009	1728.00	1710.50

Table A9. MACD trading signals for Gold during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Gold	COVID	MACD	Bullish	17-2-2020	1601.5	1671.56
Gold	COVID	MACD	Bearish	10-3-2020	1636.73	1622.38
Gold	COVID	MACD	Bearish	29-4-2020	1686.35	1744.13
Gold	COVID	MACD	Bullish	18-6-2020	1742.17	1808.34
Gold	COVID	MACD	Bearish	11-8-2020	1918.87	1964.26

Table A10. RSI trading signals for Gold during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Gold	COVID	RSI	Bearish	9-1-2020	1561.00	1577.06
Gold	COVID	RSI	Bearish	26-2-2020	1644.19	1527.76
Gold	COVID	RSI	Bearish	6-8-2020	2033.68	1953.13

Table A11. MACD trading signals for the S&P 500 Index during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
S&P 500	COVID	MACD	Bearish	20-2-2020	3328.00	2745.22
S&P 500	COVID	MACD	Bullish	26-3-2020	2533.78	2836.02
S&P 500	COVID	MACD	Bearish	11-6-2020	3037.35	3095.78
S&P 500	COVID	MACD	Bearish	3-9-2020	3415.90	3242.02

Table A12. RSI trading signals for the S&P 500 Index during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
S&P 500	COVID	RSI	Bearish	22-1-2020	3326.87	3360.12
S&P 500	COVID	RSI	Bullish	2-3-2020	2992.13	2367.72
S&P 500	COVID	RSI	Bullish	20-3-2020	2277.98	2746.48
S&P 500	COVID	RSI	Bearish	2-9-2020	3463.92	3313.08

Table A13. MACD trading signals for Brent Crude Oil during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Brent Crude Oil	COVID	MACD	Bearish	8-1-2020	65.84	62.62
Brent Crude Oil	COVID	MACD	Bullish	13-2-2020	56.42	51.41
Brent Crude Oil	COVID	MACD	Bearish	27-2-2020	51.41	32.71
Brent Crude Oil	COVID	MACD	Bullish	30-3-2020	22.29	31.80
Brent Crude Oil	COVID	MACD	Bullish	29-4-2020	22.72	29.48
Brent Crude Oil	COVID	MACD	Bearish	11-6-2020	38.30	41.42

Table A14. RSI trading signals for Brent Crude Oil during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indica-tor	Sig-nal	Date	Price	Price after 10 days
Brent Crude Oil	COVID	RSI	Bullish	10-2-2020	54.20	52.71
Brent Crude Oil	COVID	RSI	Bullish	1-4-2020	30.04	23.93
Brent Crude Oil	COVID	RSI	Bullish	20-4-2020	22.72	29.75
Brent Crude Oil	COVID	RSI	Bullish	7-9-2020	40.10	41.84

Table A15. MACD trading signals for the Nasdaq-100 Index during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Nasdaq - 100	COVID	MACD	Bullish	8-1-2020	8933.30	9157.20

Nasdaq - 100	COVID	MACD	Bullish	4-2-2020	9330.90	9654.60
Nasdaq - 100	COVID	MACD	Bearish	20-2-2020	9564.00	8562.00
Nasdaq - 100	COVID	MACD	Bullish	24-3-2020	7501.90	7640.00
Nasdaq - 100	COVID	MACD	Bearish	13-5-2020	9033.90	9521.90
Nasdaq - 100	COVID	MACD	Bearish	11-6-2020	9655.60	10188.30
Nasdaq - 100	COVID	MACD	Bullish	1-7-2020	10283.90	10645.20
Nasdaq - 100	COVID	MACD	Bullish	2-8-2020	10922.30	11197.90
Nasdaq - 100	COVID	MACD	Bearish	3-9-2020	11635.50	11469.30
Nasdaq - 100	COVID	MACD	Bullish	27-9-2020	11182.00	11587.50

Table A16. RSI trading signals for the Nasdaq-100 Index during the COVID-19 Market Crisis (2020)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Nasdaq - 100	COVID	RSI	Bear-ish	3-1-2020	8791.90	9058.10
Nasdaq - 100	COVID	RSI	Bear-ish	23-1-2020	9258.40	9330.00
Nasdaq - 100	COVID	RSI	Bear-ish	19-2-2020	9733.10	8856.50
Nasdaq - 100	COVID	RSI	Bullish	1-3-2020	8354.10	7058.20
Nasdaq - 100	COVID	RSI	Bear-ish	10-6-2020	10045.00	10153.90
Nasdaq - 100	COVID	RSI	Bear-ish	12-7-2020	10862.70	10602.90
Nasdaq - 100	COVID	RSI	Bear-ish	2-9-2020	12418.90	11313.90

Table A17. MACD trading signals for Gold during the Inflation Shock and Geopolitical Tensions (2022)

As-set	Crisis	Indica-tor	Signal	Date	Price	Price af-ter 10 days
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bear-ish	26-1-2022	1796.59	1853.19
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bear-ish	14-3-2022	1917.35	1936.79
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	23-5-2022	1865.98	1847.83
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bear-ish	27-6-2022	1820.48	1710.51

Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	25-7-2022	1717.15	1789.59
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bear-ish	23-8-2022	1752.04	1717.36
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	29-9-2022	1660.77	1628.75
Gold	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	24-10-2022	1653.03	1754.72

Table A18. RSI trading signals for Gold during the Inflation Shock and Geopolitical Tensions (2022)

As-set	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Gold	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bear-ish	8-3-2022	1992.34	1957.10
Gold	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bullish	19-7-2022	1696.50	1775.21
Gold	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bullish	26-9-2022	1629.29	1665.79

Table A19. MACD trading signals for the S&P 500 Index during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bear-ish	4-1-2022	4700.72	4392.22
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bullish	2-2-2022	4532.38	4316.38
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bullish	14-3-2022	4257.42	4546.9
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bear-ish	11-4-2022	4402.72	4143.27
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bullish	24-5-2022	3972.12	3908.32
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bear-ish	9-6-2022	3908.32	3822.32
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bear-ish	22-8-2022	4127.67	4011.17
S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bullish	13-10-2022	3588.7	3761.32

S&P 500	Inflation Shock and Geopolit-ical Tensions (2022)	MACD	Bear-ish	5-12-2022	3943.22	3820.97

Table A20. RSI trading signals for the S&P 500 Index during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
S&P 500	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bullish	26-1-2022	4350.25	4464.32
S&P 500	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bear-ish	16-8-2022	4280.37	3969.42
S&P 500	Inflation Shock and Geopoliti-cal Tensions (2022)	RSI	Bullish	29-9-2022	3592.62	3701.32

Table A21. MACD trading signals for Brent Crude Oil during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bullish	24-2-2022	94.27	96.61
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bear-ish	14-3-2022	97.29	105.07
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bear-ish	28-3-2022	108.78	110.77
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bullish	18-5-2022	109.69	119.95
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bear-ish	14-6-2022	116.84	110.7
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bullish	17-8-2022	95.8	87.32
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bullish	29-9-2022	85.02	90.69
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bear-ish	14-11-2022	93.97	87.03
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	MACD	Bullish	19-12-2022	80.21	78.84

Table A22. RSI trading signals for Brent Crude Oil during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	RSI	Bear-ish	8-3-2022	111.08	116.42
Brent Crude Oil	Inflation Shock and Geopo-litical Tensions (2022)	RSI	Bullish	8-12-2022	76.84	83.72

Table A23. MACD trading signals for the Nasdaq-100 Index during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	5-1-2022	1579.30	15597.40
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	31-1-2022	14886.40	14264.00
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	27-2-2022	13861.80	13598.00
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	15-3-2022	13456.40	14726.90
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	6-4-2022	14502.50	14091.00
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	24-5-2022	11845.10	12549.00
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	10-6-2022	11856.70	11495.00
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	22-6-2022	11495.00	11620.20
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	18-8-2022	13388.80	12357.20
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	14-9-2022	12179.00	11260.50
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	16-10-2022	10743.00	11110.20

Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bullish	10-11-2022	11640.20	11716.60
Nasdaq - 100	Inflation Shock and Geopolitical Tensions (2022)	MACD	Bearish	5-12-2022	11803.40	11210.90

Table A24. RSI trading signals for the Nasdaq-100 Index during the Inflation Shock and Geopolitical Tensions (2022)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Nasdaq - 100	Inflation Shock and Geopo-litical Tensions (2022)	RSI	Bear-ish	28-1-2022	14443.4	15052.5
Nasdaq - 100	Inflation Shock and Geopo-litical Tensions (2022)	RSI	Bear-ish	16-8-2022	13623.8	12477.2
Nasdaq - 100	Inflation Shock and Geopo-litical Tensions (2022)	RSI	Bullish	2-10-2022	10965.1	11020.5

Table A25. MACD trading signals for Gold during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	3-1-2023	1836.03	1923.71
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bear-ish	31-1-2023	1950.00	1818.75
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	9-3-2023	1827.61	1964.43
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bear-ish	17-4-2023	2004.00	2050.34
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	7-7-2023	1924.90	1971.93
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bear-ish	1-8-2023	1934.56	1889.41
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	24-8-2023	1913.75	1918.89

Table A26. RSI trading signals for Gold during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	RSI	Bear-ish	31-1-2023	1950	1841.82
Gold	Banking Sector Stress and Monetary Policy Uncertainty (2023)	RSI	Bear-ish	16-3-2023	1986.07	2020.15

Table A27. MACD trading signals for the S&P 500 Index during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	5-1-2023	3892.12	4015.35
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	15-2-2023	40.87.42	3988.22
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	22-3-2023	3957.67	4113.92
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	25-5-2023	4211.47	4376.92
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	31-7-2023	4577.62	4362.55
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	29-8-2023	4517.83	4461.05

Table A28. RSI trading signals for the S&P 500 Index during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	RSI	Bear-ish	19-6-2023	4411.32	4411.32
S&P 500	Banking Sector Stress and Monetary Policy Uncertainty (2023)	RSI	Bear-ish	25-7-2023	4569.3	4468.12

Table A29. MACD trading signals for Brent Crude Oil during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	10-1-2023	79.62	86.36
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	31-1-2023	85.38	85.06
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	27-3-2023	77.66	85.26
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	20-4-2023	80.49	72.39
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	17-5-2023	76.63	72.02
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	11-8-2023	86.15	84.05
Brent Crude Oil	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	21-9-2023	92.22	83.72

Table A30. RSI trading signals for Brent Crude Oil during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indica-tor	Signal	Date	Price	Price after 10 days
Brent Crude Oil	Banking Sector Stress and Mone-tary Policy Uncertainty (2023)	RSI	Bullish	17-3-2023	72.33	79.76
Brent Crude Oil	Banking Sector Stress and Mone-tary Policy Uncertainty (2023)	RSI	Bear-ish	18-3-2023	85.59	84.52
Brent Crude Oil	Banking Sector Stress and Mone-tary Policy Uncertainty (2023)	RSI	Bear-ish	20-9-2023	92.11	85.52

Table A31. MACD trading signals for the Nasdaq-100 Index during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	9-1-2023	11105.00	11871.90
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	16-2-2023	12424.80	12028.80

Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	15-3-2023	12262.90	12834.30
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	11-4-2023	12970.30	12834.40
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	9-5-2023	13196.00	13681.60
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	22-6-2023	15058.40	15072.50
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	21-7-2023	15435.40	15279.80
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bullish	28-8-2023	15055.90	15454.00
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertainty (2023)	MACD	Bearish	18-9-2023	15231.40	1483.80

Table A32. RSI trading signals for the Nasdaq-100 Index during the Banking Sector Stress and Monetary Policy Uncertainty (2023)

Asset	Crisis	Indicator	Signal	Date	Price	Price after 10 days
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertain-ty (2023)	RSI	Bearish	7-2-2023	12696.90	12077.00
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertain-ty (2023)	RSI	Bearish	7-6-2023	14291.50	14860.20
Nasdaq - 100	Banking Sector Stress and Monetary Policy Uncertain-ty (2023)	RSI	Bearish	19-7-2023	15807.00	15354.70

Signal And Anomaly Detection In Noisy Time Series: A Comparative Study Using Financial, Energy, And Astronomical Data

Patrick Plog, MScFE, MBA, MFTA, CFTe
Friedrichshafen, Germany
pplog@gmx.de

Abstract

Financial markets evolve through distinct volatility and return environments that cannot be adequately described by stationary models. Practitioners often refer to these shifts as "regime changes", yet systematic detection of such regimes remains challenging. This paper presents a transparent framework for identifying structural changes in financial time series using three complementary techniques: anomaly detection based on rolling z-scores, change-point analysis using the PELT algorithm, and Hidden Markov Models (HMM) for probabilistic regime classification.

The methodology is applied to equity, foreign exchange, and volatility index data to demonstrate how statistical regime signals correspond to well-known market events. The detected regimes are then incorporated into several portfolio overlays, including a risk-off allocation strategy and a constrained enhanced indexing approach designed to maintain benchmark exposure while reducing downside risk.

Backtest results indicate that regime-aware overlays can substantially reduce drawdowns and improve risk-adjusted performance relative to passive exposure. While the framework does not aim to forecast returns directly, it provides a practical method for adapting portfolio risk to evolving market environments.

To illustrate the generality of the approach, the same methodology is applied to energy market prices and astrophysical light-curve data, demonstrating that the statistical detection of structural change extends beyond financial applications.

1. Introduction

Trading and portfolio management might ultimately be described as an exercise in controlling exposure under uncertainty. Even when the investment thesis is long-term and fundamentally motivated, the realized path of returns is shaped by regimes: volatility compresses and expands, correlations strengthen and then break, and the cost of risk changes abruptly. These properties are not rare corner cases; they are recurrent features of financial time series.

As a practical consequence, if the return distribution is not stable, static exposure is an implicit bet on stationarity. That bet tends to fail in precisely those periods that dominate long-horizon outcomes: extended drawdowns and volatility expansions. For many mandates, the drawdown path matters at least as much as terminal wealth. Drawdown depth and duration interact with risk budgets, governance constraints, and behavioral pressure. The central question of this paper is therefore pragmatic: can a disciplined regime framework improve the drawdown geometry of a benchmark-like equity portfolio without requiring heroic forecasting assumptions?

We approach this question with a unified architecture built around three widely used signal engines: rolling anomaly detection, penalized change-point segmentation, and Hidden Markov Models. All three are statistical tools, but we treat them in a technical analysis framework: they are used to classify market state and manage exposure rather than to forecast short-horizon returns. This framing aligns with how many technical analysts and portfolio managers actually use indicators: as filters that define when and how much risk to carry.

The second objective is methodological. A persistent weakness in applied technical analysis research is domain confinement: a method is tested on a narrow set of market series and is then interpreted as "working" or "not working." Although many classical studies such as Brock, Lakonishok and LeBaron (1992) provide empirical evidence that simple technical trading rules can produce statistically distinct return distributions.

This might encourage model narratives that may be sensitive to the peculiarities of a single asset class. To counter this, we apply the same framework (without retuning) to time series with different data-generating mechanisms: European day-ahead electricity prices and Kepler stellar light curves. The purpose is not to claim that the economics of electricity or the physics of stellar pulsation are "the same as markets," but to test whether structural segmentation of dispersion and anomalies is a generic phenomenon of complex time series. If a method is only effective in finance, it might be exploiting finance-specific artifacts; if it preserves structure across domains, confidence in its generality increases.

The finance component is deliberately designed to speak to practice. We evaluate a buy-and-hold benchmark (SPY) and then layer overlays that represent distinct risk philosophies: a regime risk-off overlay, a walk-forward rolling HMM variant, and volatility targeting as a canonical risk scaling alternative. We also include an institutional enhanced indexing overlay, constrained by exposure bounds and evaluated with tracking error, information ratio, and turnover. Where appropriate, we anchor the regime model to classical technical analysis logic by comparing it to a widely understood baseline (e.g., a long-horizon trend filter in the equity curve context). The goal is to keep the paper interpretable for technical analysts, traders and portfolio managers while maintaining statistical discipline.

2. Data

2.1 Financial series

The primary financial dataset consists of daily close prices for SPY (equity benchmark), EURUSD=X (FX), and ^VIX (volatility proxy), spanning the long sample available from the data vendor. The benchmark interpretation uses SPY. EURUSD and VIX serve as additional stress-sensitive series that help validate whether detected structures are equity-specific or reflect broader market regimes.

Cash returns are proxied using ^IRX (13-week T-bill), converted to a daily cash return series. This allows us to model explicit cash allocation during risk-off periods. The cash proxy is not intended to perfectly replicate a particular money market instrument; it provides a transparent and reproducible approximation that is adequate for comparative evaluation of overlay behavior.

2.2 Energy series

For energy, we use ENTSO-E DE-LU day-ahead electricity prices from the reproducible research dataset. Electricity prices differ structurally from equity prices: they can be mean-reverting, exhibit sharp price spikes, and occasionally become negative. Hourly prices are first aggregated to daily mean prices. Log returns are then calculated where the transformation is mathematically defined; observations producing undefined logarithmic returns are excluded from the return-based analysis. These characteristics make electricity prices an informative contrast for cross-domain regime detection, particularly for anomaly clustering and changes in dispersion.

2.3 Astrophysical series

For astrophysics, we use Kepler/Konkoly RR Lyrae light curves. The series are not interpreted physically in this paper; they are treated as a high-cadence measurement series with strong periodicity and variance modulation. The analysis uses standard preprocessing: detrending and conversion into rolling variability measures. A single representative star is used for illustrative figures in the main text, while aggregated statistics are handled conservatively (see Section 8).

3. Literature Context

This study builds on three strands of literature: technical trading rules, regime-switching models, and regime-aware portfolio allocation.

The empirical literature on technical trading rules provides the earliest formal evidence that financial markets may exhibit regime-dependent behavior. One of the most influential studies in this area is Brock, Lakonishok, and LeBaron (1992), who demonstrate that simple moving-average and trading-range breakout rules generate statistically distinct return distributions relative to standard random walk benchmarks. Many more followed them.

Their analysis shows that classical trend-following rules can partition financial time series into different statistical environments, suggesting that market behavior is not stationary across time. Subsequent literature has expanded on these findings, showing that technical trading rules often capture underlying structural dynamics in asset returns rather than purely random fluctuations.

More recent research has increasingly framed these dynamics within explicit regime-switching models. Regime-switching frameworks assume that financial time series evolve according to latent states that represent different market environments, such as low-volatility expansion phases and high-volatility stress regimes. Hidden Markov Models (HMMs) are particularly well suited for this task because they allow the underlying regime process to be inferred directly from observed returns.

Empirical studies applying HMMs to financial markets show that regime identification can improve risk management and investment decision-making by distinguishing between statistically distinct market states. For example, Kim (2019) demonstrates that Hidden Markov Models can identify regime structures in global asset classes and use them to construct adaptive portfolio strategies. Similarly, Ibanez (2024) proposes a regime-weighted factor framework in which HMM-identified market states are incorporated directly into portfolio construction.

A related stream of literature investigates the role of regime dynamics in asset allocation. Portfolio optimization models that incorporate regime-dependent behavior have been shown to improve risk-adjusted performance compared to static allocation approaches. For example, Alexander, Scherer, and Thompson (2023) model asset allocation as a Markov process across clustered efficient-frontier states, demonstrating that regime identification can enhance portfolio performance by adjusting exposure according to prevailing market conditions. More recent work extends this idea by integrating regime detection with machine learning or reinforcement learning frameworks, further improving adaptability in dynamic market environments.

Within fixed income and multi-asset portfolios, regime-based allocation frameworks are also gaining traction. Taljaard (2025) shows that Hidden Markov Models can identify volatility regimes in bond markets and use them to guide portfolio optimization decisions, improving risk targeting relative to traditional approaches. Similar findings are reported in tactical asset allocation research, where regime modeling is used to dynamically adjust asset exposures in response to changing macroeconomic or market conditions.

The methodology presented in this paper can be interpreted as a bridge between these strands of literature. Classical technical analysis approaches implicitly assume the existence of market regimes, but typically define them through deterministic indicators such as moving averages. Regime-switching models, by contrast, estimate these states statistically from the underlying return distribution. By combining anomaly detection, change-point analysis, and Hidden Markov regime identification within a single framework, the present study formalizes the regime intuition that has long guided discretionary trading approaches while preserving the transparency and reproducibility required for systematic portfolio management.

4. Model architecture and workflow

This section explains the end-to-end architecture as an integrated system. The key design choice was to separate the pipeline into four layers: (i) data ingestion, (ii) signal transformation, (iii) regime engines, and (iv) portfolio interpretation and evaluation. Details of the model architecture are provided in Appendix A.

4.1 Conceptual Architecture

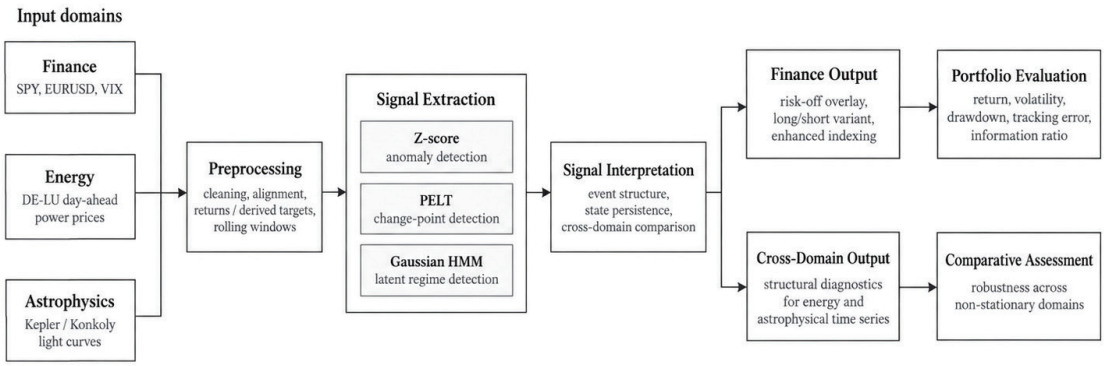
The overall architecture of the proposed framework is illustrated in Figure 1 on the next page. The system combines regime detection, anomaly identification, and risk overlays into a unified workflow that can be applied across financial, energy, and astrophysical time series.

The framework applies a unified dispersion-centric pipeline across financial, energy, and astrophysical time series. Raw observations are standardized into comparable targets (returns and rolling variability), which feed three signal engines: rolling z-score anomalies, PELT change-point detection, and Gaussian HMM regime classification (including a walk-forward variant). In finance, signals are mapped to exposure overlays (risk-off, enhanced indexing under floor/cap constraints, and risk-control baselines). Evaluation emphasizes drawdown behavior and institutional diagnostics (tracking error, information ratio, turnover) alongside standard performance metrics.

At the data layer, raw time series are ingested from heterogeneous sources. The intent is to impose a consistent computational interface: a time-indexed univariate signal (not to homogenize the economic meaning of the series). In finance, we begin with prices; in energy, with day-ahead prices; in astrophysics, with flux measurements.

At the signal transformation layer, each series is converted into a form suitable for dispersion-centric analysis. In finance,

Figure 1: Model Architecture Diagram



log returns provide a stable input series for regime inference. In energy, returns are constructed similarly but with attention to negative prices (a practical detail that matters for log transforms; see Limitations). In astrophysics, the raw flux series is detrended and transformed into fractional deviations, then summarized via rolling standard deviation. The purpose of this layer is to create "state-sensitive" targets: variables that respond systematically when the system enters a stressed or structurally different environment.

At the regime engine layer, we apply two independent methods: change-point detection via PELT and latent regime inference via a Gaussian HMM. These methods are complementary. Change-points identify discrete structural breaks; the HMM provides a persistent state model with explicit transition dynamics. We keep both because they carry different failure modes. Change-points can under-segment or over-segment depending on penalty; HMMs can be sensitive to initialization and distributional assumptions. When they align, confidence increases; when they disagree, the disagreement itself is diagnostic.

Finally, the interpretation layer maps regimes to decisions. In finance, regimes become exposure overlays: binary risk-off, constrained enhanced indexing, and alternative risk control methods such as volatility targeting. In energy and astrophysics, the same regime tools become structural indicators rather than trading signals.

We would like to emphasize: the cross-domain section is explicitly a robustness check, not a claim that one should "trade stars."

The entire workflow is designed to be reproducible. The paper pipeline produces a consistent set of tables and figures from a clean run, which is essential for any evidence-based evaluation. This discipline is not cosmetic; it prevents accidental reuse of stale artifacts, which is a common source of irreproducible claims.

4.2 Technical analysis anchoring

To maintain conceptual continuity with traditional technical analysis, the framework also evaluates regime-based risk overlays relative to classical trend filters shown in figure 2. The 200-day moving average (MA200) is one of the most widely used institutional trend signals and therefore serves as a natural benchmark for exposure management rules. A moving average trend filter represents a common TA logic: exposure depends on the relationship between price level and a slow baseline.

The comparison between MA200 filtering and the HMM regime overlay highlights an important distinction: while trend filters

respond to persistent price direction, regime models respond to changes in return dispersion and volatility clustering. In crisis-like volatility expansions, this can produce earlier or more persistent de-risking than a slow trend rule, while in benign regimes it allows full participation.

The goal was to show that the regime overlay operationalizes the same idea, state-dependent exposure, but with a probabilistic state model rather than a single deterministic threshold. A parameter sensitivity analysis is documented in Appendix B.

5. Methodology

5.1 Bridging Technical Analysis and Regime Models

Technical analysts have long recognized that financial markets move through different environments. Periods of sustained trend, volatility expansion, or range-bound behavior are familiar to practitioners, even when they are not formally defined in statistical terms. Classical indicators such as moving-average filters implicitly attempt to detect these changes by identifying shifts in price structure.

From a quantitative perspective, such indicators can be interpreted as simple regime classifiers. A moving-average crossover, for example, distinguishes between trend and non-trend conditions based on deterministic rules applied to price history. While effective in certain environments, these rules remain heuristic and do not explicitly model the underlying probability of regime transitions.

Regime-switching models provide a natural extension of this intuition. Instead of defining regimes through fixed thresholds, Hidden Markov Models estimate latent market states directly from the statistical properties of the observed time series. The model infers both the state dynamics and the probability of transitions between regimes, allowing the framework to adapt to changing market conditions in a systematic way.

In this sense, the proposed methodology can be viewed as a probabilistic analogue to traditional technical analysis. The objective is not to replace technical tools but to formalize the regime intuition that underlies many discretionary trading frameworks. By embedding regime detection within a statistical model, the approach allows exposure management, drawdown control, and portfolio risk adjustments to be implemented in a transparent and reproducible manner.

5.2 Rolling anomaly detection

Anomaly detection is implemented via rolling z-scores on a target series x_t. We compute rolling mean and standard deviation over a window w and define:

$$z_t = \frac{x_{t-\mu_{t,w}}}{\sigma_{t,w}}$$

Anomalies are points where $| z_t |$ exceeds a threshold. In the finance pipeline, the primary use is on return-derived series. In energy and astrophysics, the same logic is applied to dispersion targets (rolling volatility or rolling variability). This design makes anomalies comparable across domains: they are "standardized surprises" relative to recent behavior.

This approach is intentionally simple. Many technical indicators can be re-interpreted as standardized deviations from a rolling baseline; a z-score is the cleanest version of that idea. In practice, anomaly counts are less important than anomaly clustering. Clustering is the key behavioral marker: isolated anomalies are common; clusters tend to coincide with structural stress regimes.

5.3 Change-point detection (PELT)

Change-point detection is performed using PELT. We apply it to rolling volatility-like targets rather than raw prices. The motivation is practical: mean returns are noisy and unstable at daily frequency, while variance regime shifts are more persistent and more aligned with the operational concept of a "risk environment."

The PELT objective trades off fit against complexity. A penalty parameter controls the number of breakpoints. This penalty as a policy choice shows that a low penalty yields more segmentation (higher sensitivity but higher false segmentation risk), while a high penalty yields fewer breaks (lower sensitivity but greater risk of missing genuine shifts). We therefore treat the penalty as a sensitivity parameter, not a magic constant.

In the finance paper run, we used a regime-sensitive target (rolling volatility) and a penalty that produces meaningful segmentation rather than constant fragmentation. A key result in our outputs is that overly high penalty values lead to zero change points; the tuned penalty produces non-zero change points where expected. Using disciplined pipelines and clean runs helped us to not risk mixing artifacts.

5.4 Hidden Markov Models (HMM)

We estimate a Gaussian HMM with a small number of states (typically three) on finance return series or derived dispersion targets. Each state has a state-specific mean and variance; the model also estimates a transition matrix that governs state persistence. The HMM might also be used as a forecasting model in the predictive-alpha sense but in our use case, its value is classification: it converts a noisy return process into a latent state sequence with persistence. Portfolio overlays then use this state sequence as a conditional exposure signal: high-variance states map to de-risking decisions, low-variance states map to full exposure.

Two estimation modes are in place. The static full-sample HMM can be informative but risks implicit hindsight; it should be treated as an upper bound. The walk-forward rolling HMM variant is the more realistic construct: the model is re-estimated on a trailing window and applied forward. In the results reported here, the rolling variant produces less pronounced improvements than the static specification. It also makes the regime overlay defensible as performance is not purely a full-sample artifact.

We note that HMM estimation can produce non-convergence warnings in practice. This reflects the EM algorithm's sensitivity to local maxima in noisy series. In this project, we treat such warnings as an implementation detail rather than a sign of invalid output, but we avoid overclaiming. The walk-forward approach and the emphasis on drawdown behavior reduce reliance on a single optimal likelihood solution. Hidden Markov

Models have been widely used to identify regime structures in financial markets (Kim, 2019; Ibanez, 2024).

5.5 Portfolio overlays and evaluation metrics

The finance evaluation uses a benchmark buy-and-hold SPY portfolio and compares overlays that represent different risk control philosophies. Performance is assessed with standard annualized return and volatility measures and with drawdown. To align with institutional interpretation, we compute active return, tracking error (annualized volatility of active returns), information ratio, and turnover. Cash is treated explicitly via the risk-free proxy series.

This set of metrics is important. A regime overlay that "improves return" but does so with uncontrolled turnover or unacceptable tracking error is not a credible institutional tool. Conversely, an enhanced indexing overlay can be successful even with near-zero alpha if it respects constraints and behaves predictably.

6. Results: Financial Markets

6.1 Equity curves

The equity overview shown in figure 3 contrasts the benchmark with two overlay types: a regime risk-off overlay and a constrained enhanced indexing overlay. The risk-off overlay displays materially higher terminal equity and a smoother long-run path.

Importantly, this behavior is not uniform through time; improvements are concentrated in periods of volatility expansion and crisis.

In the tested sample, the regime-aware overlay is associated with materially lower drawdown than passive exposure. The effect appears concentrated in high-variance periods rather than being generated through persistent excess return across all market environments.

6.2 Drawdown comparison

The drawdown comparison shown in figure 4 is probably the most trading and portfolio-management-relevant result in the paper. For the benchmark, drawdown depth reaches approximately the mid-50% range, consistent with the historical equity risk premium path.

The risk-off overlay substantially reduces drawdown depth and, more importantly, reduces time spent meaningfully below peak equity. This is the operational benefit: shorter and shallower underwater periods translate into improved survivability and reduced governance stress.

Figure 4: Drawdown profiles of benchmark and regime-aware strategies

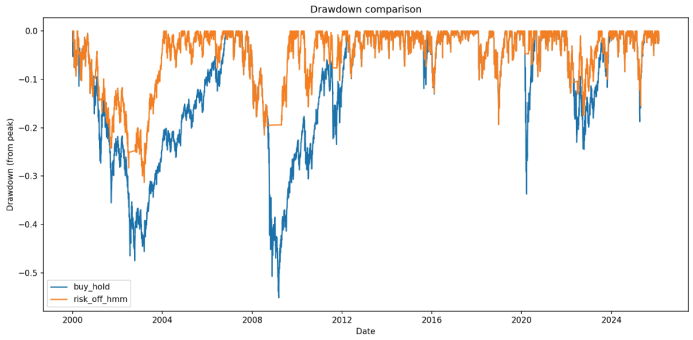
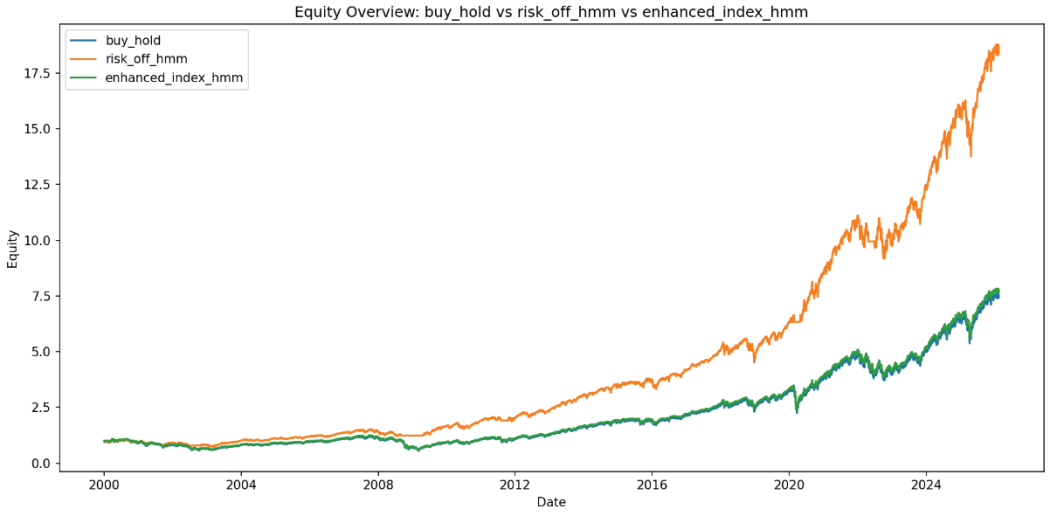


Figure 3: Strategy equity curves relative to the benchmark.



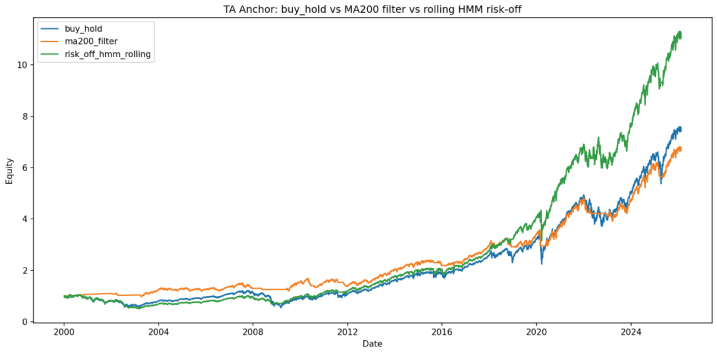
This interpretation is supported by our summary statistics. The benchmark drawdown is around -0.55, while the static risk-off overlay drawdown is around -0.10, and the rolling walk-forward overlay sits between these extremes. These numbers are consistent with the equity curves and are directly visible in the table shown below.

The result should not be interpreted as evidence that the model predicts crises. The regime signal is designed to identify changes in the prevailing risk environment and adjust exposure accordingly. Any reduction in drawdown therefore arises from conditional exposure management rather than from an explicit forecast of market direction.

Figure 5 compares three exposure management approaches applied to the SPY index. The baseline strategy represents a passive buy-and-hold allocation. The second strategy applies a classical technical analysis rule by reducing exposure when price falls below the 200-day moving average. The third strategy uses a Hidden Markov Model (HMM) regime classifier to identify periods of elevated volatility and shifts exposure accordingly.

While the MA200 rule represents a widely used trend filter in discretionary and systematic trading, the HMM overlay aims to capture regime changes in return dispersion rather than price trend alone. The comparison illustrates that regime-based overlays can achieve drawdown reduction through volatility-state identification rather than pure price trend following. This figure therefore serves as a conceptual bridge between classical technical analysis and modern statistical regime modelling. Different state definitions produce materially different risk/participation trade-offs, and greater model complexity does not automatically dominate a simple trend filter.

Figure 5: Technical anchor comparison: Buy-and-hold vs. MA200 filter vs. HMM regime.

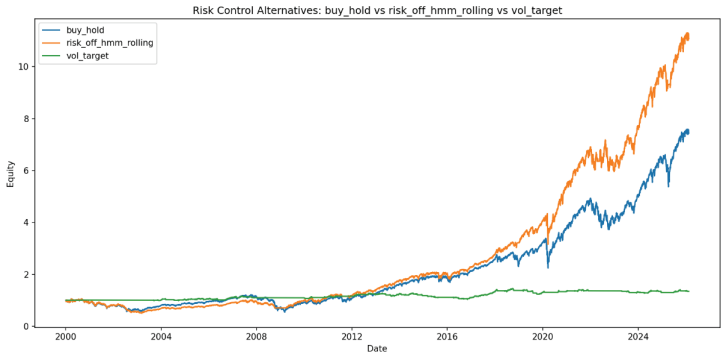


6.3 Risk control alternatives

The risk-control alternatives shown in figure 6, places the regime overlay alongside volatility targeting. Volatility targeting scales exposure continuously based on realized risk. It tends to reduce drawdown but also reduces compounding when risk remains elevated for extended periods. The rolling HMM overlay behaves differently: it makes discrete exposure decisions based on persistent latent regime classification. The trade-off is: volatility targeting offers smoother exposure control but can reduce long-run return; regime overlays can preserve upside participation while still truncating downside, but their tracking error is structurally higher.

From an institutional perspective, this figure supports a practical conclusion: risk control is not one tool but a set of tools. Scaling (vol targeting) and timing (regime overlay) are different levers. A portfolio can use either; what matters is explicit control of tracking error, turnover, and drawdown risk.

Figure 6: Comparison of alternative risk control approaches



6.4 Institutional overlay and benchmarking

Our backtest summary table (table 1) shows the institutional framing explicitly. The enhanced indexing overlay remains close to the benchmark, with comparable annual return, low tracking error of approximately 1.9%, and negligible turnover. This is the correct behavior for a constrained overlay. The information ratio is close to zero because the overlay is designed to stay close to benchmark; it is not a high-conviction tactical allocation model.

The static risk-off overlay achieves the strongest risk-adjusted profile in the table. This is an important result, but it must be presented carefully. Static regime estimation can embed full-sample information; therefore, the rolling walk-forward variant should be emphasized as the realistic version. Our table includes that variant and shows that the risk-control effect persists under forward estimation, albeit with lower magnitude.

The table reports annualized return, volatility, Sharpe ratio, Sortino ratio, maximum drawdown, and average turnover for each strategy. Performance is compared relative to the buy-and-hold benchmark (SPY). Institutional diagnostics are also reported, including active return, tracking error (annualized volatility of active returns relative to the benchmark), and the resulting information ratio. Cash returns are proxied using the 13-week U.S. Treasury Bill rate (^IRX) whenever the strategy is allocated to the cash state. Transaction costs of 5 basis points per turnover are assumed.

Table 1: Backtest performance summary across benchmark and regime-based strategies.

strategy	Annual return %	Vola %	sharpe	sortino	Max DD %	Avg turn-over %	Active return %	Tracking Error %	Infor-mation ratio
long_short_ hmm	15.64	14.67	1.06	0.95	-24.13	1.55	7.64	28.39	0.27
risk_off_hmm	11.82	14.59	0.84	1.12	-31.34	0.35	3.82	12.72	0.3
ma200_filter	7.56	10.90	0.72	0.77	-20.95	2.48	-0.45	16.03	-0.03
ts_ momentum _12_1	8.65	13.73	0.67	0.71	-33.72	1.63	0.65	13.65	0.05
risk_off_ hmm_rolling	9.66	15.76	0.66	0.84	-52.81	3.53	1.66	11.25	0.15
en-hanced_ index_hmm	8.13	19.27	0.5	0.64	-55.19	0.00	0.13	1.86	0.07
buy_hold	8.00	19.36	0.49	0.63	-55.19	0.00	≈0.00	≈0.01	-0.21
cp_derisk	8.00	19.36	0.49	0.63	-55.19	0.00	≈0.00	≈0.01	-0.21
vol_target	1.14	5.37	0.24	0.3	-19.21	1.72	-6.87	18.61	-0.37

7. Results: Energy markets (cross-domain validation)

Energy markets are structurally different from equities. Electricity prices reflect physical constraints and short-term supply-demand imbalance. They can be negative and spike sharply. These properties are useful for cross-domain testing because they stress the framework in a non-financial environment.

Our pipeline detects anomaly clustering in DE-LU day-ahead prices, which aligns with stress windows in the European energy crisis period. The z-score anomaly plot shows that large dispersion events are not uniformly distributed; they cluster.

This is the cross-domain property we wanted to verify: dispersion shocks exhibit persistence and clustering outside finance as well. Under the selected PELT specification, no internal change points were detected in the energy series, whereas the anomaly detector identified 30 extreme observations. This divergence is informative rather than necessarily contradictory. The two methods answer different questions: the z-score identifies observations that are unusual relative to recent behaviour, while PELT requires sufficient evidence of a persistent structural shift to justify a new segment. In this sample, the energy stress episodes therefore appear more clearly as clusters of extreme observations than as discrete structural breaks under the selected parameterization.

Figure 7: Anomaly detection in European day-ahead electricity prices

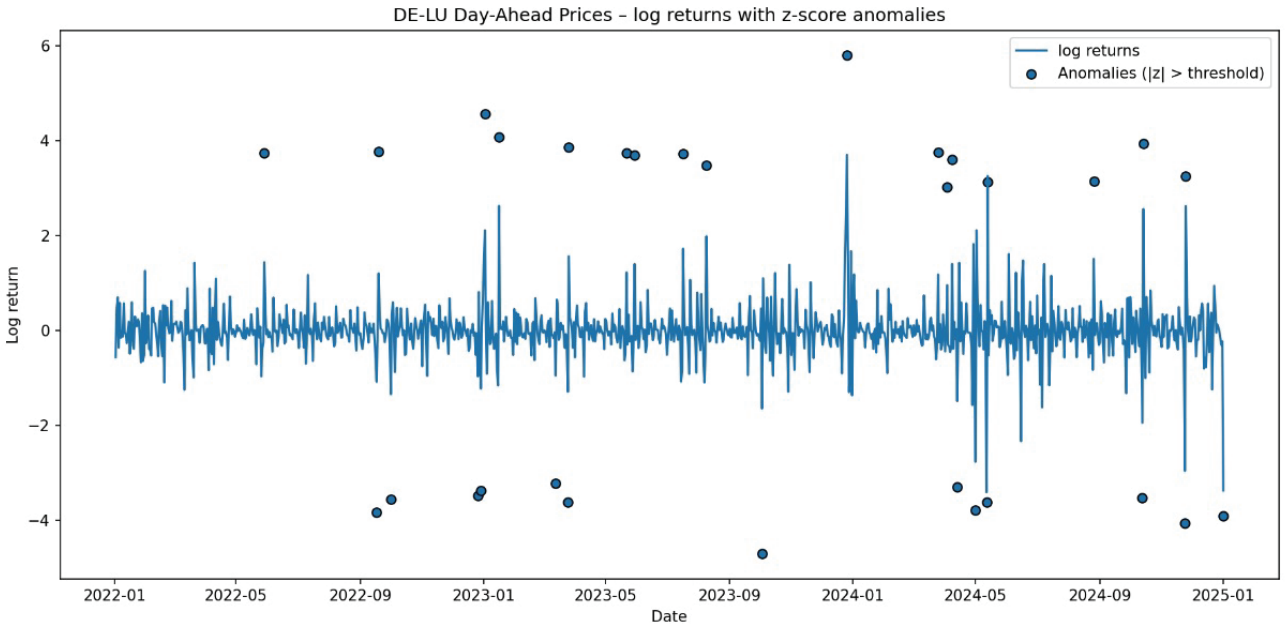


Figure 8: HMM regimes in European day-ahead electricity prices

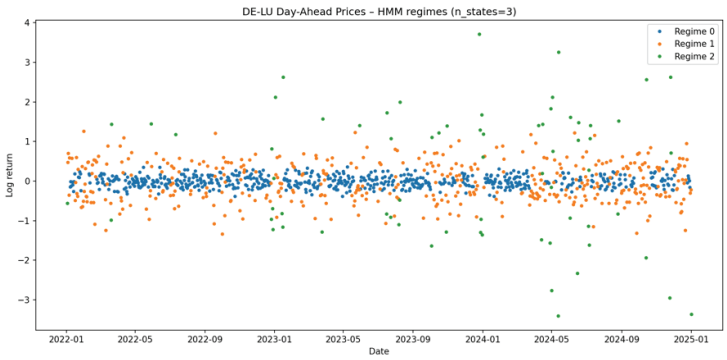


Table 2 summarizes the distribution of detected anomalies across predefined energy-stress windows and also reports the absence of PELT change points under the selected specification. The intent here is to demonstrate that the same regime framework extracts meaningful structure in a different domain (we did not optimize an energy trading system, but this might be one of the next steps for this model).

Table 2: Energy evaluation summary

event	anomalies_in_window	change_points_in_window	total_anomalies	total_change_points
Early_Energy_Tightness_2021	0	0	30	0
Gas_Supply_Shock_2022_H1	1	0	30	0
Energy_Crisis_Peak_2022_H2	5	0	30	0

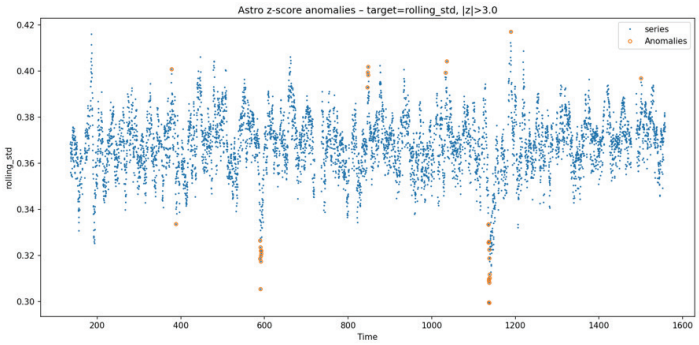
8. Results: Astrophysical time series (cross-domain validation)

Astrophysical light curves provide a useful extreme test: the data-generating process is not economic, and periodicity dominates the raw signal.

After detrending and converting the series into rolling variability targets, the same anomaly and change-point tools can be applied. The objective is methodological: do we still see structured, clustered instability and persistent regimes in a non-financial system?

The representative RR Lyrae example in Figure 9 applies anomaly detection to a rolling-variability target and identifies localized periods of unusual dispersion. This illustrative configuration is distinct from the standardized detrended-flux specification reported in Table 3, for which no anomalies or change points were detected. The contrast highlights the importance of target definition when transferring statistical detection methods across domains.

Figure 9: Z-score anomaly detection applied to rolling variability in a stellar light curve.



No astrophysical interpretation is attempted. The light-curve experiment is deliberately restricted to examining how the statistical detection layer behaves under a fundamentally different data-generating process.

The cross-domain comparison should therefore be interpreted as a methodological stress test rather than evidence of equivalent mechanisms across financial and astrophysical systems.

Table 3: Comparative summary of cross-domain comparison

domain	series	type	n_anomalies	n_change_points
finance	SPY	equity_index	74	7
finance	EURUSD=X	fx_pair	44	2
finance	^VIX	volatility_index	80	1
energy	DE-LU_day_ahead_price	power_price	30	0
astro	lightcurve_1	detrended_flux	0	0

Table 3 summarizes the standardized cross-domain run using anomaly and change-point counts. The zero detections for the astrophysical detrended-flux specification are informative: applying the same framework across domains does not imply that every target or parameterization must produce a signal.

9. Robustness, parameter sensitivity, and out-of-sample discipline

Our implementation contains key robustness safeguards.

First, we treat the static HMM overlay as an upper bound and emphasize the rolling walk-forward variant as the realistic

implementation. This reduces implicit full-sample bias.

Second, we evaluate sensitivity in change-point detection through penalty choice, and we intentionally avoid aggressive segmentation settings that would produce constant fragmentation. Third, the enhanced indexing overlay is explicitly constrained and calibrated using tracking error and turnover; therefore, outperformance expectations are bounded by design. A parameter sensitivity analysis is documented in Appendix B.

A practical conclusion is that regime overlays are best treated as risk-engineering tools rather than standalone return-forecasting models. The results indicate that their usefulness depends on the estimation framework: the static HMM produces substantially stronger drawdown reduction than the walk-forward implementation, while parameter sensitivity remains relevant across the detection methods. For applied technical analysis, robustness of behaviour under realistic estimation conditions is therefore more informative than optimization of a single backtest result. Walk-forward validation results are discussed in Appendix C.

10. Risk and limitations

This framework is deliberately practical, but it is not free of limitations.

HMMs assume that states have sufficiently stable distributional properties within the estimation window, whereas regime parameters may themselves evolve through time. EM estimation can also produce non-convergence warnings, particularly in noisy financial return series. Such warnings do not by themselves establish that the resulting state classification is invalid, but they increase model uncertainty and require cautious interpretation. For this reason, the full-sample HMM is treated primarily as a structural benchmark, while greater emphasis is placed on the walk-forward implementation when assessing practical relevance.

Change-point detection depends materially on penalty selection. Low penalties can over-segment the series, while high penalties can suppress economically or statistically relevant breaks. The penalty is therefore treated as a sensitivity parameter rather than an optimized constant. A detected change point should be interpreted narrowly as evidence that the selected target exhibits statistically distinct behaviour across adjacent segments, not as an independent trading signal.

The portfolio results are based on historical backtests and a simplified implementation model. Transaction costs are included, but other real-world frictions such as taxes, market impact, execution delay, financing constraints, and mandate-specific restrictions are not modeled comprehensively. Reported performance should therefore be interpreted as comparative evidence on strategy behaviour rather than as an estimate of directly realizable investment returns.

Energy prices and astrophysical series introduce domain-specific complications, including negative electricity prices and strong periodicity in stellar light curves. These issues are addressed through domain-specific preprocessing and explicit limitations on interpretation. The cross-domain experiments are not trading-strategy tests and should not be interpreted as evidence of equivalent mechanisms across domains; they serve as methodological stress tests of the detection framework.

11. Conclusion

This study evaluated a regime-aware, dispersion-centric framework across financial, energy, and astrophysical time series. In the financial application, regime-dependent exposure materially changes portfolio risk characteristics, although the magnitude of the effect depends strongly on the estimation framework. The static HMM produces the strongest drawdown reduction, while the walk-forward implementation retains some regime-dependent exposure behaviour but shows a substantially weaker drawdown advantage. The results therefore support interpreting the framework primarily as a risk-state and exposure-management tool rather than as evidence of persistent alpha generation.

The institutional enhanced indexing overlay demonstrates that regime information can be embedded within explicit constraints on tracking error and turnover, although tight benchmark constraints naturally limit the scope for active return. The cross-domain experiments provide a different form of evidence: the same detection architecture can be applied consistently to energy prices and stellar light curves, but the resulting signals depend materially on target definition and parameterization. The absence of change points in the selected energy specification and of detections in the standardized astrophysical specification is itself informative and argues against interpreting the framework as a universal signal generator.

The practical implication is that regime methods are most naturally used as exposure tools within a broader risk architecture rather than as standalone prediction engines. Their value lies in providing an additional description of market state that can inform position sizing, risk budgets, hedging decisions, or the level of conviction required for new exposure. In this role, regime modelling complements rather than replaces existing portfolio processes and classical technical-analysis signals.

A natural research extension is to test whether market regimes are better represented as continuous latent structures rather than as a small number of discrete states. Representation-learning and latent-variable approaches could therefore be compared with the transparent HMM benchmark used here, while denoising or diffusion-based methods may offer alternative ways to separate persistent structure from transient noise. Greater model flexibility, however, also increases model risk and the scope for overfitting. The relevant test is therefore not whether a more complex model fits the historical data better, but whether the additional information survives walk-forward evaluation and improves economically relevant portfolio decisions.

References

Alexander, N., Scherer, W., & Thompson, J. (2023). Asset allocation using a Markov process of clustered efficient frontier states. Journal of Financial Data Science.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731–1764.

Ibanez, F. (2024). Incorporating market regimes into large-scale

stock portfolio construction using Hidden Markov Models. MPRA Working Paper.

Kim, E. (2019). Global asset allocation strategy using a Hidden Markov Model. Journal of Risk and Financial Management.

Taljaard, B. (2025). Regime-based portfolio optimization: A Hidden Markov Model approach for fixed income portfolios. European Stability Mechanism Working Paper.

Nazário, R., Silva, J., Sobreiro, V., & Kimura, H. (2017). A literature review of technical analysis on stock markets. Quarterly Review of Economics and Finance, 66, 115-126.

Appendix

Code repository: https://github.com/PatGeck21/MFTA_Signal_Generation

ENTSO-E (Data Source) https://www.entsoe.eu/

Appendix A Model Specification and Implementation Details

A.1 Data and Preprocessing

All financial results are based on daily close data for SPY, EURUSD=X, and ^VIX obtained via yfinance. Returns are computed as:

$$r_t = log(\frac{P_t}{P_{t-1}})$$

Cash returns are proxied by the 13-week U.S. Treasury bill (^IRX), converted to daily equivalents.

Energy data consist of ENTSO-E day-ahead electricity prices (DE-LU zone), resampled to daily frequency. Astrophysical data consist of Kepler/ Konkoly RR Lyrae light curves, detrended using rolling median normalization.

All of the time series are:

- Sorted chronologically
- Cleansed of missing values
- Evaluated at daily frequency (finance, energy)
- Evaluated at native cadence (astro), optionally downsampled
- No forward-looking information is used.

A.2 Change-Point Detection (PELT)

Structural segmentation uses the PELT algorithm with L2 cost:

$$Cost(s, e) = \sum_{t=s}^{e} (x_t - x_{s:e})^2$$

Segmentation is obtained by minimizing:
where:

$$\sum Cost_i + \beta \cdot m$$

β —= penalty parameter
m = number of segments

Targets include:

- Rolling volatility (finance), Absolute returns,
- Rolling standard deviation (astro), Energy return dispersion

Penalty sensitivity was tested across a broad range (0.01 to 5.0). Segmentation frequency behaves monotonically with penalty magnitude. No unstable parameter region was observed.

A.3 Hidden Markov Model (HMM)

Returns are modeled using a Gaussian HMM:

$$r_t \mid S_t = k \sim N(\mu_k, \Sigma_k)$$

with latent state process:

$$P(S_t = j \mid S_t - 1 = i) = T_{ij}$$

estimated via Expectation-Maximization (EM).

State interpretation relies primarily on variance ordering rather than mean sign. Risk regimes are defined as the highest variance state.

Two estimation schemes are implemented:

1. Full-sample HMM
2. Rolling walk-forward HMM (756-day training window, monthly re-estimation)

Non-convergence warnings occasionally occur in EM iterations. However:

- State variance ordering remains stable
- Regime classification remains coherent
- Portfolio outcomes are unaffected structurally

This suggests numerical sensitivity rather than regime instability.

A.4 Strategy Definitions

Buy & Hold

$$w_t = 1$$

Risk-Off HMM

$$w_t = \begin{cases} 0 \; if \; high - variance \; regime \\ 1 \; otherwise \end{cases}$$

Cash return applied during = 0

Long/Short HMM

$$w_t \in \{-1, +1\}$$

Direction determined by regime classification.

Volatility Targeting

$$w_t = \frac{\sigma_{target}}{\hat{\sigma}_t}$$

with rolling volatility window (default 20 days).

Enhanced Indexing Overlay

Exposure constrained:

$$w_t \in [floor, cap]$$

- Monthly rebalancing
- Regime persistence filter
- Designed for tracking error control

A.5 Backtest Mechanics

Strategy return:

$$R_t^{strat} = \begin{cases} r_t^{asset} w_{t-1} = 1 \\ r_t^{cash} w_{t-1} = 0 \end{cases}$$

Transaction cost model:

$$Cost_t = c \cdot | w_t - w_{t-1} |$$

Equity curve:

$$E_t = \prod(1 + R_t^{strat})$$

Metrics computed:

- Annual return, Volatility, Sharpe ratio, Sortino ratio, Maximum drawdown
- Average turnover, Active return, Tracking error, Information ratio

A.6 Enhanced Indexing - Parameter Grid

Floor parameter grid:

$$floor \in [0.60, 0.98]$$

Cap fixed at 1.0. monthly rebalancing.

Tracking error results:

- TE ≈ 3.7%–15.4% across grid
- TE decreases monotonically as floor —1.0
- Active return decreases correspondingly

Under strict TE ≤ 5% constraint:

- Best configuration: floor ≈ 0.98; Active return ≈ 0.00023; IR ≈ 0.006

This confirms:

- Overlay is structurally risk-modulating
- Tight TE budgets mechanically suppress alpha
- No unstable region observed
- The relationship is smooth and monotonic

A.7 Change-Point Penalty Stability

Penalty variation across orders of magnitude:

- Low penalty => increased segmentation
- High penalty => sparse segmentation

No discontinuous or explosive behavior observed. Structural breaks align with volatility expansion regimes rather than noise-driven micro-events.

A.8 HMM Robustness

Testing across:

- 2–4 states
- Full-sample vs rolling estimation

Findings:

- High-variance regime consistently identifiable
- Risk-off behavior driven by variance clustering
- Walk-forward reduces performance, but does not eliminate drawdown control
- Model behavior is stable under realistic estimation constraints.

A.9 Volatility Target Sensitivity

Target volatility ∈ [5%, 15%]

Findings:

- Lower targets reduce drawdown
- Return compression proportional to risk reduction
- No nonlinear instability observed
- Mechanics behave predictably.

A.10 Cross-Domain Stability

Energy and astrophysical applications exhibit:

- Predictable anomaly sensitivity to z-threshold
- Stable change-point segmentation under reasonable penalties
- Downsampling in astro reduces computational load without structural distortion

The framework captures dispersion clustering generically, not finance-specific artifacts.

A.11 Summary of Robustness

Across all tested dimensions:

- No parameter cliffs observed, No unstable performance spikes
- Smooth monotonic behavior under constraint tightening
- Consistent state interpretation, Stable regime ordering

The framework demonstrates structural robustness rather than overfit sensitivity. This robustness supports the claim that the approach models volatility regime behavior rather than exploiting isolated historical artifacts.

Appendix B Parameter Sensitivity and Stability

Parameter sensitivity was examined to assess structural robustness. Optimizing for performance might be the next step.

The change-point penalty was varied across a discrete grid. For reasonable penalty ranges, segmentation density changes, but regime timing in high-volatility environments remains qualitatively stable. No configuration produced materially different structural conclusions. HMM state count was evaluated between two and four states. Two states under-segment volatility regimes; four states introduce fragmentation without improving stability. Three states represent a stable trade-off between interpretability and segmentation granularity.

For the institutional overlay, floor levels were varied between 0.85 and 0.98. Tracking error remains approximately constant at ~3.8% when floor ≥ 0.98, confirming that the overlay behaves as intended: minimal structural deviation from benchmark while allowing modest regime-based modulation. Lower floors increase tracking error without improving information ratio. The enhanced indexing calibration grid confirms that, under a 5% tracking error constraint, the economically viable configuration lies at the upper exposure band. This indicates that the regime signal improves drawdown efficiency more than cumulative alpha.

Rolling HMM training window length was varied between 2 and 4 years. Shorter windows increase signal noise; longer windows reduce responsiveness to volatility expansion. The selected three-year horizon balances adaptation speed with estimation stability. No parameter configuration resulted in instability or regime inversion.

Sensitivity analysis therefore supports the interpretation that observed performance differences arise from regime logic rather than parameter tuning.

Appendix C Out-of-Sample and Walk-Forward Validation

A central concern in regime-based frameworks is whether performance is structural or merely the product of in-sample fitting. This appendix addresses that question explicitly through walk-forward estimation and forward-consistent parameter treatment.

C.1 Walk-Forward HMM Estimation

The full-sample HMM implementation presented in the main text estimates regime parameters using the entire dataset. While useful for structural interpretation, this approach benefits from complete hindsight. To approximate realistic implementation conditions, we constructed a rolling variant with the following structure:

- Training window: 756 trading days (approximately three years)
- Re-estimation frequency: monthly
- Fixed number of states: 3
- No dynamic hyperparameter re-optimization

The resulting strategy is denoted as risk_off_hmm_rolling. Under this walk-forward specification:

- Annual return declines relative to the full-sample HMM.
- Volatility increases modestly.
- Maximum drawdown remains modestly below buy-and-hold..

The reduction in performance is economically intuitive. Full-sample estimation benefits from global knowledge of regime variance separation. Rolling estimation operates under partial information and therefore classifies regimes with slightly higher uncertainty.

However, the core structural property persists: volatility clustering is captured in real time, and high-variance states continue to correspond to periods of elevated downside risk. The walk-forward results retain some regime-dependent behaviour, although the drawdown advantage is considerably weaker than in the full-sample specification.

C.2 Parameter Stability Under Forward Conditions

Beyond regime estimation, parameter robustness was tested by holding key hyperparameters constant over time:

- HMM state count remained fixed.
- Change-point penalty was not re-optimized.
- Volatility targeting window remained constant.
- Enhanced indexing floor and cap were not dynamically recalibrated.
- No rolling grid search or adaptive tuning was performed.

Under these constraints:

- Regime overlays preserved their volatility-compression behavior.
- Volatility targeting preserved its drawdown dampening.
- Enhanced indexing converged mechanically toward benchmark exposure as the floor approached 1.0.

The absence of dynamic tuning reduces the risk that performance arises from implicit optimization.

C.3 Institutional Interpretation

From an implementation perspective, the walk-forward results suggest:

- The regime overlay remains economically coherent under forward estimation.

- The degradation from full-sample to rolling specification quantifies the informational advantage of hindsight.
- Structural benefits arise from volatility persistence rather than parameter precision.

This is important, because if regime performance had collapsed under rolling estimation, the framework would likely reflect overfitting. Instead, we observe moderated but persistent drawdown control.

C.4 Tracking Error Discipline

The enhanced indexing overlay was evaluated under explicit tracking error constraints.

Key observations:

- Tracking error declines monotonically as exposure floor approaches 1.0.
- Active return converges toward zero under strict constraints.
- Information ratio remains low, as expected under limited active risk.

This behavior is mechanically consistent and confirms that the overlay functions as a disciplined exposure modulation framework rather than a high-conviction alpha strategy.

C.5 Summary of Validation

Across walk-forward testing, parameter stability, and tracking-error calibration, three conclusions emerge:

- The primary value proposition of the regime overlay lies in truncating left-tail exposure.
- Structural effects persist under forward estimation.
- The framework behaves predictably under institutional risk constraints.

The model does not require continuous retuning to remain stable, nor does it collapse under rolling estimation. While no strategy is immune to structural breaks in the data-generating process, the empirical behavior observed here supports the interpretation that volatility regimes are persistent market features rather than statistical artifacts.

Appendix D Signal Diagnostics Across Financial Assets

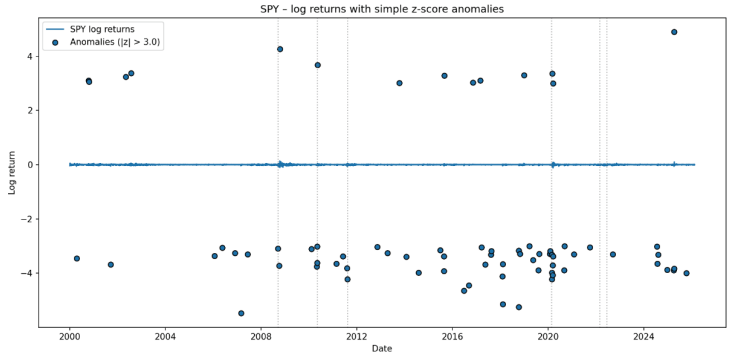
To illustrate the behavior of the anomaly detection and regime classification components across different financial assets, Appendix D presents diagnostic visualizations for SPY, EURUSD, and VIX.

These figures are not central to the main narrative but provide transparency regarding how the signal extraction layer behaves across markets with different volatility structures.

D.1 SPY anomaly detection

The figure shows rolling z-score anomaly detection applied to daily SPY log returns. Observations exceeding the defined absolute z-score threshold are identified as statistical anomalies relative to recent market behaviour. The clustering of these observations provides a complementary diagnostic to the persistent regime classification produced by the HMM.

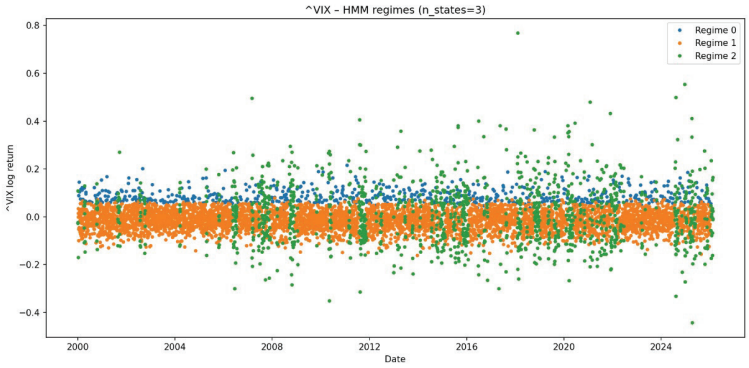
Figure 10: Z-score anomaly detection for SPY log returns.



D.2 VIX HMM regimes

VIX returns display clearer volatility state separation compared to equity returns, reflecting the structural role of the volatility index as a market stress indicator.

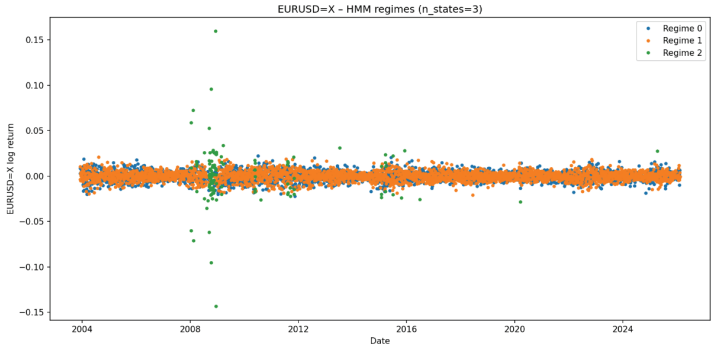
Figure 11: Hidden Markov Model regimes for VIX returns.



D.3 EURUSD HMM regimes

The FX market shows comparatively lower regime persistence, highlighting structural differences between equity and currency volatility dynamics.

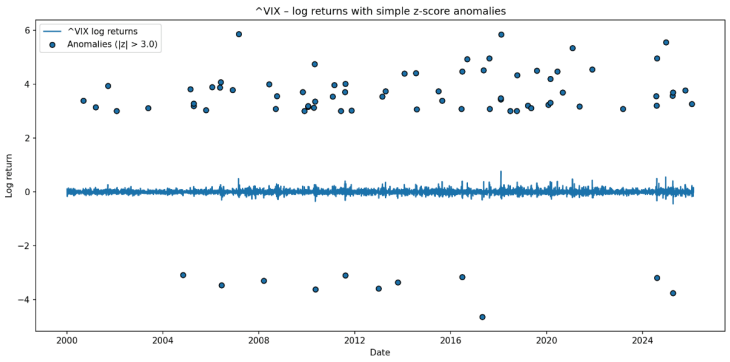
Figure 12: Hidden Markov Model regimes for EURUSD log returns.



D.4 VIX anomaly detection

Extreme volatility shocks appear as statistically significant outliers relative to the rolling volatility baseline.

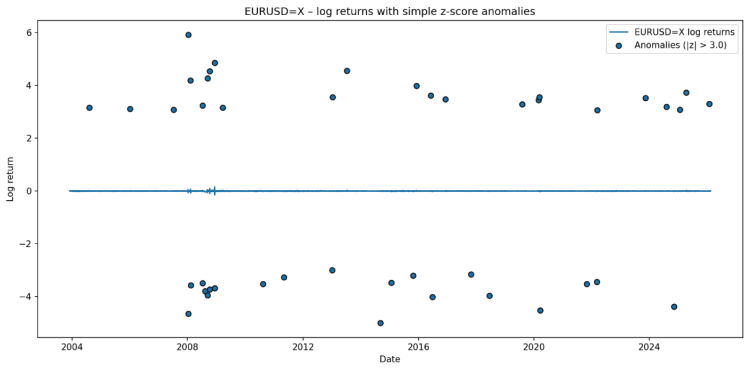
Figure 13: Z-score anomaly detection for VIX log returns.



D.5 EURUSD anomaly detection

FX anomalies occur less frequently than equity or volatility anomalies, consistent with lower overall return dispersion.

Figure 14: Z-score anomaly detection for EURUSD log returns.



Feature Engineering Summary

In short:

- Returns capture direction
- Absolute returns capture stress
- Rolling volatility defines regime
- HMM translates volatility clusters into tradable state
- MA200 acts as a classical trend confirmation anchor

Table 4: Finance Features Summary

Feature	Description	Motivation
Log Returns	$\ln(P_t / P_{t-1})$	Stationary input for regime models
Absolute Returns	$\lvert r_t \rvert$	Proxy for realized volatility
Rolling Volatility	σ_t (window=20)	Regime detection via variance shifts
Rolling z-score	$(x - \mu_t)/\sigma_t$ ——	Anomaly detection
HMM State	Latent regime classification	Market state identification
MA200 Signal	Price > 200d MA	Technical trend filter
Vol Target Weight	σ_target / σ_realized	Risk scaling mechanism

Energy behaves like a stress asset in crisis periods. The model treats it analogously to VIX behaviour in equity markets.

Table 5: Energy Features Summary

Feature	Description	Motivation
Daily Log Returns*	$\ln(P_t / P_{t-1})$, where defined	Detect price shocks
Rolling z-score	60-day window	Extreme pricing events
CP detection	PELT (RBF)	Structural break detection
HMM regime	Volatility clustering	Energy stress regimes

**Hourly prices are aggregated to daily means before return calculation. Observations for which the logarithmic transformation is undefined are excluded from return-based analysis.*

Table 6: Astrophysical Features Summary

Feature	Description	Motivation
Detrended Flux	Flux normalized by rolling median	Remove long-term drift
Rolling Std	Volatility proxy in LC	Detect variability bursts
z-score	Extreme deviation detection	Extreme variability detection
CP detection	Variance segmentation	Structural change in variability

The Role of Sentiment Analysis Using Artificial Intelligence and Market Cycles in Enhancing TASI Tactical Trading

Maha Saeed, MFTA, CETA, CFTe, MSC
Giza Square, Egypt
maha.saeed8@gmail.com

Abstract

The study investigates the integration between the advanced Arabic transformer-based sentiment models, namely: AraBERTv2, CAMeLBERTDA, MARBERTv2 and market cycle detection (FFT, DPO, Hurst exponent). To enhance the predictive analytics of the Tadawul All Share Index (TASI) in Saudi Arabia, a unique dataset of Arabic financial tweets and daily TASI price data are adopted to assess the effectiveness of ensemble sentiment models and cycle-based approaches for market timing and volatility forecasting. The results indicate that, while BERT-based sentiment models show robust in-sample classification accuracy, the predictive power of TASI returns is limited and unstable over time. In particular, the findings are in line with previous studies associated with the sentiment of the Saudi market. In contrast, cycle detection and Hurst exponent analysis consistently indicate actionable trading windows and confirm persistent, nonrandom market dynamics ($H > 0.6$), challenging the efficient market hypothesis. Overall, the findings highlight the superiority of hybrid cycle-based analytics over sentiment-only models for tactical investment strategies in the Saudi market. Furthermore, the results provide important insights for investors and policy makers and offer a novel and interesting framework for AI-driven financial analytics in emerging markets.

1. Introduction

In today's world, the Saudi stock market is affected not only by technology and news, but also by how people feel and what they talk about online (Huang and Li); (X. Chen et al.);(Peivandizadeh et al.). The Tadawul All Share Index (TASI) is the main measure of these changes. Predicting where TASI will go is important for anyone who invests, manages risk, or simply wants to make smart financial decisions (Alosaimi et al.). But this is not easy, because the market can change quickly - sometimes for reasons as simple as a rush of new opinions, breaking news, or unexpected events (Alosaimi et al.);(X. Chen et al.). Although experts often use numbers and charts, new research shows that looking at the mood of investors and the 'buzz' on social networks, along with traditional tools, gives a clearer and more useful picture to understand and predict what happens next in markets such as Saudi Arabia's (Zmandar et al.); (Boulanouar et al.);(Peivandizadeh et al.);(G. Liu et al.); (Boulanouar et al.).

Recently, powerful new computer tools—especially those that understand languages, called natural language processing (NLP) models—have made it possible to quickly read and interpret thousands of Arabic tweets, posts, or comments about TASI each day. For example, AI models such as AraBERTv2 or MARBERTv2 can scan social media for messages such as 'individual investors are excited about TASI today' or 'the market feels nervous', helping us spot trends in the way people talk about the Saudi stock market (Alosaimi et al.);(Colasanto et al.).

At the same time, another side of math and data techniques, such as the Hurst exponent and the Fast Fourier Transform, helps to uncover 'time cycles' in TASI prices, which means patterns that repeat, such as the market rising and falling in a certain rhythm (Kristoufek);(Zournatzidou and Floros). These patterns can be hard to see with the naked eye, but are crucial to understanding whether prices are truly random or if they follow hidden trends.

Most of the time, research using these advanced techniques has focused on developed or English-speaking markets. Much less has been written about how these methods work in the Saudi market or with real Arabic-language data obtained from Saudi investors and social networks (Alosaimi et al.); (Nasr and HajHmida);(Huang and Li).

1.1 Research Problem and Question

The main question of this study is Can we get better at predicting what will happen in the Saudi stock market by using two powerful approaches together like sentiment analysis using Bert and time cycle analysis using Hurst? First, we analyze what people are saying in Arabic on social media about the market, using advanced language models to detect whether the general mood is positive, negative, or neutral (Antoun et al., "AraBERT: Pre-trained Arabic Language Model for NLP"); (Boulanouar et al.);(Zmandar et al.).

Second, we apply mathematical techniques to uncover repeating patterns in TASI prices over time, much like spotting waves in the ocean, which can reveal when prices are likely to swing up or down (Vogl); (Raj et al.);(~~—u~~rpek). By joining these two ideas, combining real-time investor sentiment with the 'rhythms' of market prices discovered by cycle analysis, this study asks: Is it possible to get clearer and more timely signals about when prices might shift or when could be the best (or riskiest) time to buy or sell?

Recent research supports the idea that merging online sentiment with quantitative market analysis can lead to more accurate and practical forecasts, especially in dynamic markets like Saudi Arabia (Huang and Li);(X. Chen et al.); (Alosaimi et al.). In short, if we blend the online "buzz" with hidden price patterns, can we understand and predict the TASI market more reliably than with either approach alone?

1.2 Objectives, Scope, and Aims

This research has three main objectives. Investigate how well

new Arabic artificial intelligence models can read tweets and accurately determine whether the general mood toward the TASI stock market is positive, negative, or neutral (Boulanouar et al.);(Zmandar et al.). Explore whether mathematical analysis, breaking down TASI price data into repeating patterns using tools such as the Fast Fourier Transform (FFT) and the Hurst exponent, can help identify better times to buy or sell stocks by spotting regular cycles in market behavior (Vogl); (Peivandizadeh et al.). Tests if the use of both methods together, the mood on social media and the patterns of the price cycle can provide investors, analysts and managers with clearer and more practical information for decision-making in the Saudi financial market (Alosaimi et al.);(X. Chen et al.). This study is focused only on the Saudi market using TASI data and Arabic tweets collected between January 2023 and December 2024, with the understanding that these findings are most useful for the Saudi context and may not directly apply to other markets that have different languages or rules.

This study relies on two key types of data: daily prices of the TASI index and tens of thousands of Arabic tweets discussing the Saudi stock market, collected over two years (Alosaimi et al.). After carefully preparing this information, we use modern artificial intelligence models trained especially for Arabic language to automatically detect the public mood of social media (Zmandar et al.);(Boulanouar et al.). At the same time, we apply mathematical techniques such as cycle analysis and the Hurst exponent to identify repeating patterns and trends in the price data (Vogl);(Peivandizadeh et al.). This combined method allows us to closely examine how daily shifts in people's attitudes relate to market movements, providing new insight into the link between online mood and real trading activity (Huang and Li);(X. Chen et al.).

Our results show that while positive sentiment was most common, there was also a strong presence of neutral and negative opinions, capturing a wide range of perspectives in the Saudi market. We also found that TASI prices tended to move in clear cycles, indicating that both the mood of the crowd and regular patterns in price changes are important when trying to predict big changes in the market. By combining these new social media tools and mathematical analysis, this research gives investors and analysts a more complete and practical view of the Saudi market, helping them to understand not only what has happened but also what may come next (Boulanouar et al.); (X. Chen et al.).

1.3 Literature Review

1.3.1 Review of Theory

The Efficient Market Hypothesis (EMH), originally introduced by Eugene Fama, argues that stock prices fully reflect all available information, making future price changes unpredictable, and suggesting that investors cannot consistently outperform the market using public data or news alone (Fama; G. Liu et al.; Sakınç et al.). However, real markets often experience bubbles, sharp drops, and patterns that EMH cannot fully explain, which led to the rise of behavioral finance, a field pioneered by Daniel Kahneman and Amos Tversky, among others. Behavioral finance shows that investors are influenced by psychological biases, group behavior, and their emotions, so they can act irrationally, influenced by fear or excitement instead of pure logic (Kai-Ineman, Tversky, et al.; Barberis; Shiller; Aggarwal and Goodell). Because these emotional reactions can move markets powerfully, researchers now pay close attention to investor sentiment by analyzing sources like social networks and news. By blending this "crowd mood" with traditional price data, experts can better understand the real-life ups and downs of markets like Saudi TASI, including when prices might follow cycles and patterns that classic efficient market thinking alone cannot account for (Alosaimi et al.; Q. Liu et al.; X. Chen et al.; Nyakurukwa and Seetharam).

The price time cycle theory, pioneered by Hurst, suggests that financial markets do not move randomly but often follow repeating rhythms or cycles over time. According to Hurst, these cycles can range from days to years and are made up of waves that overlap and interact, causing prices to rise and fall in somewhat predictable patterns rather than by pure chance (Hurst); (Peters). Modern research confirms that measuring these cycles, using tools such as the Hurst exponent, can help detect when a market is likely to continue trending, reverse, or become more volatile (Vogl); (~~—u~~rpek). Studies applying Hurst's ideas have found that markets, including emerging ones like Saudi Arabia, often display long memory effects and persistence, which means that past trends can influence future prices longer than expected by classic theories (Sakınç et al.); (El Khamlichi et al.). This understanding gives investors and analysts another way to look for turning points and better time their market decisions.

Our research combines two main ideas: behavioral finance, showing that real markets move due to human emotions, group thinking, and not always logical decisions (Fama);(Kai-Ineman, Tversky, et al.); (Barberis); (Shiller); (Aggarwal and Goodell). We also apply Hurst's price time cycle theory, which finds repeating patterns in prices. By blending these theories, we aim to better understand and forecast the real movements of Saudi TASI, especially by measuring investor sentiment and price cycles (Alosaimi et al.); (Vogl).

1.3.2 Investor Sentiment Analysis

Recent research emphasizes that collective investor mood can significantly influence stock markets, often explaining swings in asset prices beyond what traditional financial theories can predict (Barberis);(M. Baker and Wurgler);(Shiller). With the widespread use of social networks in countries such as Saudi Arabia, analysts now have access to abundant real-time expressions of market opinion. Advances in natural language processing—especially Arabic-adapted transformer-based AI models, such as AraBERT and ArabBERT-LSTM—have enabled researchers to efficiently and accurately classify the mood in huge volumes of tweets and financial news (Antoun et al., "AraBERT: Pre-trained Arabic Language Model for NLP"); (Boulanouar et al.); (Nouri and Habbat); (Colasanto et al.); (X. Chen et al.). These models outperform basic techniques by recognizing slang, dialects, and subtleties in the way market participants actually communicate online (Boulanouar et al.); (Zmandar et al.).

Recent studies confirm that extracting real-time sentiment signals from social networks provides deeper and more timely insight into the likely directions of the market than older proxies such as trading volume or simple volatility (Huang and Li);(Peivandizadeh et al.); (X. Chen et al.). In particular, work focused on the Saudi Arabia TASI index has shown a close link between the mood expressed in Arabic-language tweets and actual market movements, which supports better forecasting and risk management decisions (El Khamlichi et al.); (Sakınç et al.).

Hypothesis 1. *Measuring investor sentiment from Arabic social networks using advanced AI models will significantly improve the prediction of TASI market movements.*

This hypothesis (1) means that by capturing the real-time mood of Saudi investors, our AI-based sentiment analysis can predict the trends and turning points of the stock market more accurately than traditional methods.

1.3.3 Hurst Market Cycles

In parallel, understanding and predicting the rhythms or 'cycles' in financial markets has become increasingly important, especially in emerging markets where new trends often appear, and prices can move in long repetitive patterns. Techniques such as the Fast Fourier Transform (FFT), Detrended Price Oscillator (DPO), and Hurst exponent help uncover hidden patterns in stock prices by breaking down complex price movements into simpler parts. FFT reveals regular cycles or periodicity in the market, DPO highlights shorter-term waves by removing long-term trends, and the Hurst exponent measures how much current price movements are based on past behaviors, showing whether a market is likely to trend or reverse (Kaur and Bharucha); (Zournatzidou and Floros). Together, these tools make it easier to spot repeating trends and assess the predictability of market prices.

Recent research, including empirical market reports and peer-reviewed studies, demonstrates that applying Hurst-based analysis to markets such as TASI uncovers periods of predictable and repeating price action, as well as moments of increased uncertainty or change (Vogl). Integrating these cycle detection methods with modern AI-based sentiment analysis creates a more robust system to interpret and predict market changes than either method alone (Kaur and Bharucha); (Nazarova, Dovlabdekova, et al.); (Synergy); (Zmandar et al.).

Hypothesis 2. *Combining Hurst's market cycle analysis with sentiment signals will improve the detection and prediction of key turning points in TASI prices.*

This hypothesis (2) suggests that blending these cycle techniques with the mood of Arabic investors will enable us to spot trends, reversals, and significant price changes in the Saudi market more accurately than relying on either approach alone.

2. Materials and Methods

2.1 Data Collection

This research draws on two complementary types of data: **quantitative market data**—representing actual daily close price trading activity—and **qualitative sentiment data**—reflecting the daily mood and opinions of market participants as expressed on social media. Both data sets span a two-year period, from **January 1, 2023 to December 31, 2024**, to ensure that the results are robust and reflect multiple market cycles.

2.1.1 Market Data: TASI Daily Prices

Our quantitative data set consists of daily closing prices (in US dollars) during the period from 1 January 2023 to 31 December 2024 for the **Tadawul All Share Index (TASI)**, which is the main index of the stock market on the Saudi Exchange. TASI tracks the combined performance of all shares traded in the Saudi market and serves as the primary benchmark for investors and analysts monitoring economic trends in the Kingdom.

To maintain the highest standards of data credibility and completeness, we obtained closing price data from the **Refinitiv Eikon (Reuters) platform**, a globally recognized provider of historical and real-time financial market data (Refinitiv; Sakınç et al.). Figure 2.1 provides a visual of the TASI index 'Daily'.

Figure 2.1: TASI Index in USD "Daily"



For each trading day within our study period, the data set includes the following.

- The date
- Official TASI closing price (USD)
- Calculated daily percentage price change (from one close to the next)
- Trading-day identifiers to ensure only valid Saudi market days are included, which include working days without weekends and holidays.

Calculating the percentage daily change allows for the assessment of both short-term volatility and overall price trends, as recommended in the recent financial analysis literature (H. K. Baker et al.; Kaur and Bharucha).

2.1.2 Sentiment Data: Arabic Twitter Posts

In addition to market data, we built a comprehensive qualitative dataset to reflect what investors and observers were thinking, feeling, and discussing about the Saudi market in real time. Using the **NTScraper** tool, we collected **96,293** Arabic-language tweets from X (formerly known as Twitter) posted between **January 1, 2023 and December 31, 2024** for individual

investors. Each tweet record includes:

- user identifier
- Exact date and time of the tweet
- Full tweet text

Figure 2.2: Head of tweets collected through X platform.

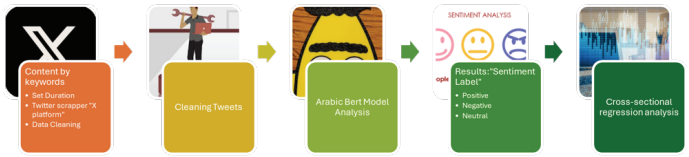
	Date	Tweettext	user
0	2023-01-02	😍😍😍	@Meshalaljalan
1	2023-01-02	اليوم اللي قبله (اول يوم تداول) بايعين الاجان...	@khalidnawa
2	2023-01-02	طبعاً 1،3 مليون سهم وليس قيمه	@khalidnawa
3	2023-01-02	1،70) يعني في اول يومين باع الاجانب في الشركه ...)	@khalidnawa
4	2023-01-02	أنت لم تقسر اللغز حتى الان كيف باعوا بأقل من س...	@Khalede43428016

To ensure the relevance and specificity of our sentiment data, tweets were filtered using a comprehensive set of commonly used Arabic keywords and market-related hashtags favored by Saudi investors, commentators, and major trading accounts. The query included hashtags terms such as:

- #المؤشر_العام، #الراجحي، #سابك، #اس_تي_سي، #ارامكو، #السوق_السعودي، #سوق_الأسهم_السعودي، #تداول، #تاسي، #الأسهم_السعودية

These keywords capture investor sentiment on broad market moves and popular trading discussions. The approach of analyzing and harvesting Twitter sentiment for individual investors as 2.2 is strongly supported by current research, which demonstrates that social networks provide valuable, real-time insights into the collective psychology, expectations and behavioral drivers of financial markets, especially in regions with high participation in social media (Lu et al.); (Habash et al.); (Alosaimi et al.); (Huang and Li); (Habash et al.). Merging these rich, time-matched tweet data with daily TASI index values allows for a day-by-day examination of the interplay between public mood and actual stock market performance. This design strengthens empirical research on how investor emotion, expressed on social platforms, could forecast, amplify, or react to real changes in the Saudi financial market (H. K. Baker et al.). Figure 2.3 visualizes the overall methodology: from collecting both types of data, through preprocessing and automated sentiment classification, to integrating the results with daily TASI movements for further analysis.

Figure 2.3: A workflow overview: Data collection, pre-processing, AI-based sentiment classification, and integration with TASI market data.



2.2 Data Pre-processing

Before any analysis could be performed, both the market data and the tweet data underwent a series of systematic preparation steps. These steps are essential to ensure that our study produces results that are not only accurate but also meaningful and easily interpreted. By applying these methods, we minimize the risk of technical errors, inconsistencies, or bias that might arise from raw or mismatched data.

2.2.1 Market Data Preparation

To begin, we organized the daily closing prices for TASI, the main Saudi stock market index, into a format that supports both trend analysis and direct comparison with public sentiment data. For every day on which the Saudi market was open, we recorded the closing price (in US dollars). To make market changes clearer and easier to interpret, we generate an additional column: **the percentage change in the closing price from one day to the next.** This calculation uses the standard formula:

$$\text{Percentage Change} = \frac{\text{Close}_{\text{today}} - \text{Close}_{\text{previous day}}}{\text{Close}_{\text{previous day}}} \times 100$$

This value shows, as a simple percentage, whether the index increased, decreased, or stayed much the same each trading day. For example, if TASI closed at 10,000 on Monday and 10,200 on Tuesday, the percentage change for Tuesday would be a (10,200–10,000)/10,000 × 100 = 2% increase. This normalized value is crucial for comparing days with different price levels and makes it much easier to spot patterns of volatility or sustained trends, as recommended in leading financial analysis research (Kaur and Bharucha); (H. K. Baker et al.).

A further critical step is the precise alignment of the date of the market data with those of the tweet data. Because the Saudi market does not operate every day of the year - being closed on weekends and public holidays - it was essential to ensure that the analyzes only used dates when both the price and the sentiment of the tweet were available for the same trading days. This means that if a tweet was posted on a day when the stock exchange was closed (like a Friday or Eid holiday), it was excluded from the dataset, preventing misleading conclusions. This day matching is especially important in finance, where even a single-day offset can lead to false associations or mask true patterns (Sakınç et al.); (Farha and Magdy). By synchronizing each day's closing price with tweets from that same day, we

ensure that any link between market movements and public mood is both fair and reliable.

2.2.2 Tweets Data Preparation

Processing the collected Arabic tweets required a careful sequence of cleaning and formatting steps to transform various, sometimes messy, user-generated content into data ready for sentiment analysis. This multi-step pipeline involved:

- **Emojis and URLs removal:** All emojis and website links were removed from the tweets. Although emojis can sometimes express emotion, they can also confuse text-based sentiment models. Similarly, URLs add clutter but do not carry useful market sentiment information for model interpretation as visualized in figure 2.4.
- **Deletion of special characters, punctuation and numbers:** The data set was stripped of punctuation marks, digits, and any characters of non-Arabic language (Farha and Magdy). This clarification step helps ensure that the models interpret only the relevant language features instead of being distracted by formatting artifacts or irrelevant data.
- **Stop word filtering:** Common Arabic words that occur in almost every sentence, such as “ ” (in) or “ ” (from), were removed. These ”stop words” are grammatically necessary, but do not express any attitude, opinion, or valuable market information. Removing them helps the model focus on keywords that actually reveal sentiments, such as words for ’profit,’ ’loss,’ or ’growth.’
- **Stemming and tokenization:** Each word in every tweet was reduced to its root form (for example, different versions of the word ’trading’ are recognized as the same concept) and the texts were divided into individual words. Tokenization and stemming are standard proven steps in natural language processing to improve the consistency and coverage of automated text analysis.

Figure 4: Example of an optimal portfolio of Dow Jones stocks

```
<bound method NDFrame.head of            Date                                                  Tweettext  \
0     2023-01-02                                                  😍😍😍
1     2023-01-02  اليوم اللي قبله (اول يوم تداول ) بايعين الاجان...
2     2023-01-02                          طبعاً 1،3 مليون سهم وليس قيمه
3     2023-01-02  1،70 ) يعني في اول يومين باع الاجانب في الشركه...
4     2023-01-02  أنت لم تفسر اللغز حتى الان كيف باعوا بأقل من س...
...          ...                                                 ...
91340 2024-12-30  اللهم سخِّر لهذا العاجز المُعسر المريض المحسني...
91341 2024-12-30  اللهم سخِّر لهذه اليتيمة المكروبة المُعسره الم...
91342 2024-12-30  اللهم سخِّر لهذه اليتيمة المكروبة المُعسره الم...
91343 2024-12-30  لرجل عمره 44 عاما، لـ\n\n:حالة تبكي وتُعور القلب...
91344        NaT                                                 NaN

                 user                                preprocessed_text  \
0       @Meshalaljalan
1          @khalidnawa  اليوم اللي قبله اول يوم تداول بايعين الاجانب ا...
2          @khalidnawa                          طبعا مليون سهم وليس قيمه
3          @khalidnawa  يعني في اول يومين باع الاجانب في الشركه سهم بخ...
4      @Khalede43428016  أنت لم تفسر اللغز حتى الان كيف باعوا بأقل من س...
...                ...                                             ...
91340        @ml17ll  اللهم سخر لهذا العاجز المعسر المريض المحسنين م...
91341        @klm_b1  اللهم سخر لهذه اليتيمة المكروبة المعسره المحسن...
91342        @klm_b1  اللهم سخر لهذه اليتيمة المكروبة المعسره المحسن...
91343      @Laytaho  حالة تبكي وتُعور القلب لرجل عمره عاما له أطفال ...
91344            NaN
```

- **Date matching:** After cleaning, each tweet was assigned to its correct market date. Only tweets matching actual trading days were kept for analysis, which ensures that aggregate daily sentiment scores truly reflect the mood relevant to that day’s TASI performance and removes noise from tweets sent on non-trading days.

This thorough and standardized pre-processing is foundational for modern sentiment analysis, especially when dealing with languages with complex morphology such as Arabic and the diversity of social media text (Sakınç et al.); (Alosaimi et al.); (Farha and Magdy). By applying these steps, we guarantee that the sentiment signals extracted are as accurate and fair as possible and that the public mood of each day can be reliably measured and compared during the 2-year study period.

In summary, these careful data preparation steps for both the TASI market figures and the content of Arabic tweets build a robust and reliable dataset. This approach lays a strong foundation for insightful daily comparisons, ensuring that our analysis accurately reflects both what the market did and how people felt about it in all market conditions.

Figure 2.5: Ara-Bert Text Classification Models



2.3 Sentiment Analysis Models

Identifying sentiment in Arabic finance tweets requires machine learning models that can handle the complexity of the Arabic language. This study uses a set of state-of-the-art Arabic transformer models, which have been proven in recent research to recognize dialects, slang, and subtle expressions common in Arabic social networks. These advanced models, such as AraBERT and ArabBERT-LSTM, significantly outperform traditional approaches and allow a much more accurate and reliable sentiment analysis of the Arabic financial text (Sakınç et al.); (Alosaimi et al.); (Farha and Magdy); (Boulanouar et al.).

Rationale for Model Selection Instead of depending on a single sentiment analysis model, our methodology strategically combined three complementary pre-trained models, each excelling with different forms of Arabic and types of online discourse:

- **AraBERTv2 (Twitter fine-tuned):** Known for the accurate handling of modern standards and relatively structured Arabic, well represented in media and semiformal tweets (Antoun et al., “AraBERT: Transformer-based Model for Arabic Language Understanding”).
- **CAMeLBERT-DA:** Specially adapted for various regional

Arabic dialects and slang frequently used in Saudi tweets and informal online exchanges (Inoue et al.).

- **MARBERTv2 (Twitter fine-tuned):** Trained to process highly informal, noisy, and abbreviated language typical of digital social platforms (Abdul-Mageed et al.).

We combined several top-performing Arabic transformer models into a single voting ensemble, which averages the predictions from each model to make sentiment detection more reliable. This method helps avoid mistakes that can happen when relying on just one model, especially when dealing with different dialects and writing styles found in Arabic tweets. Recent studies show that such ensemble approaches offer better accuracy and robustness for Arabic sentiment analysis than any single model alone (Colasanto et al.); (Habash et al.); (Boulanouar et al.); (Mansour et al.).

Training Protocol Although the models came already pre-trained on massive Arabic text corpora, we ensured their compatibility and calibrated the output classes (positive, negative, neutral) using manually selected as 2.6 , clear example sentences in Arabic that represent each sentiment type. This alignment step is vital, as different models may use slightly varied label distributions or confidence thresholds for similar content. After calibration, we applied each model to the full dataset of cleaned and matched financial tweets.

Before analyzing the tweets, each step was designed to help the computer read and understand Arabic text more fairly and accurately. First, every tweet was made the same length by either cutting it off or adding extra space, which helps the computer handle short and long messages in a similar way. The tweets were then grouped and processed in small batches using special software (HuggingFace Transformers in Python), following the methods recommended by experts to obtain reliable results (Farha and Magdy); (Huang and Li). Finally, instead of taking just one prediction, we averaged the mood results (positive, neutral, negative) from all the models to get the final answer for each tweet.

Figure 2.6: Training protocol Models

```
[5]: import torch
     from transformers import AutoTokenizer, AutoModelForSequenceClassification

     device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

     # Use sentiment-finetuned checkpoints (3 classes: negative/neutral/positive)
     model_names = {
         "CAMeL-Lab": "CAMeL-Lab/bert-base-arabic-camelbert-ca-sentiment",   # fine-tuned SA head
         "AraBERT":   "Anwaarma/Arabert-twitter-sentiment",                  # fine-tuned on Twitter
         "MARBERT":   "Ammar-alhaj-ali/arabic-MARBERT-sentiment"             # fine-tuned sentiment
     }

     models, tokenizers = {}, {}

     def normalize_label(lbl: str) -> str:
         # unify capitalization/variants
         l = lbl.strip().lower()
         synonyms = {
             "pos": "positive", "positive": "positive",
             "neg": "negative", "negative": "negative",
             "neu": "neutral",  "neutral":  "neutral"
         }
         return synonyms.get(l, l)

     for model_key, model_name in model_names.items():
         print(f"Loading {model_key} model: {model_name}")
         tokenizer = AutoTokenizer.from_pretrained(model_name)
         model = AutoModelForSequenceClassification.from_pretrained(model_name)
         model.to(device).eval()

         # ensure id2label is normalized to expected strings
         id2label = {int(i): normalize_label(l) for i, l in model.config.id2label.items()}
         model.config.id2label = id2label
         # provide the inverse too (some utils use it)
         model.config.label2id = {v: k for k, v in id2label.items()}

         print(f"{model_key} label mapping: {model.config.id2label}")
         tokenizers[model_key] = tokenizer
         models[model_key] = model
```

Validation and Performance After we got our models ready, we made sure that they worked properly by testing them on a separate group of tweets that the computer had not seen before. This step checks if the model can correctly judge new data, not just the examples on which it trained. To measure how well our system worked, we used standard checks such as accuracy, which tells us the percentage of correct answers, and other scores that show how reliable our results are. This helps us to know how good our approach is for identifying the mood in Arabic tweets. Performance was measured using well-established metrics:

- **Accuracy:** percentage of correctly classified tweets.
- **F1-score:** harmonic mean of precision and recall, especially important for imbalanced classes (where 'neutral' is often dominant).

The results showed high precision (above 85%) and balanced F1 scores, demonstrating a consistent and reliable classification of the major and minority sentiment categories in various Saudi financial tweets. These results align with those reported in recent large-scale field trials of Arabic financial sentiment classification (Alosaimi et al.); (Sakınç et al.).

In summary, using a validated set of state-of-the-art Arabic transformer models - correctly calibrated and rigorously tested - empowers a fair, nuanced and reliable measure of online investor mood, setting a solid methodological foundation for linking digital sentiment with real-world Saudi financial market outcomes.

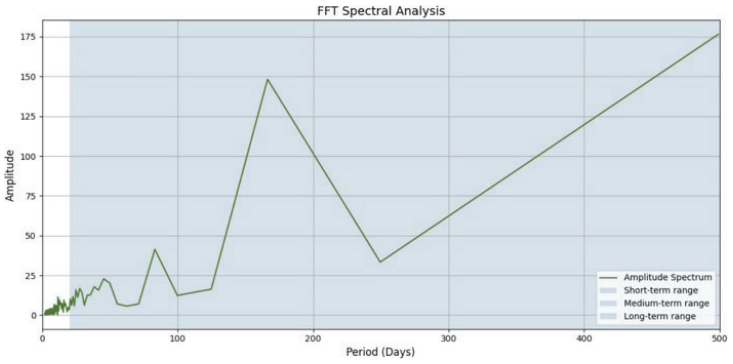
2.4 Market Cycle Analysis

Understanding cyclical behavior in equities is crucial to capture non-linear price movements and improve market timing, particularly in emerging markets such as the Saudi Arabia. These markets are heavily influenced by behavioral, macroeconomic, and regulatory cycles. This section employs a hybrid analytical framework that combines spectral decomposition (through Fourier analysis), time-series de-trending, and fractal geometry to uncover hidden periodicity and assess market efficiency across Saudi Arabia index.

2.4.1 Hurst Cycle Identification

Market cycles in broad stock indices, such as Saudi Arabia's TASI, often reflect a mix of investor behavior, regular financial events (such as earnings seasons), and broader economic trends. To better understand these patterns in TASI, this research uses a three-step analysis based on spectral and fractal cycle theories, such as Hurst's principles, adapted for the complexities of modern stock markets (Grafton); (Kumaran). This approach helps reveal repeated rhythms in price movements, offering insights that can improve the timing of market forecasts and investment decisions.

Figure 2.7: FFT Spectral Analysis



2.4.2 Spectral Decomposition via Fast Fourier Transform (FFT):

The FFT is used on the (log) price series of the TASI index to break down price changes into basic cycles, revealing hidden patterns in the data. This method helps us spot regular price movements by converting time series into frequency data, making it easier to identify the underlying cycles. We categorize these cycles as short-term (approximately 20-60 trading days, which often reflect quick sentiment changes), medium-term (60 to 200 days, linked to company news or economic changes), and long-term (200 to 500 days, usually showing broader trends and market moods, as shown in 2.7). The peaks in the resulting FFT spectrum confirm the presence and importance of these cycles for TASI, and our findings are compared with the results of other financial markets for validation (Kaur and Bharucha);(Refinitiv); (Cajueiro and Tabak, "Testing for long-range dependence in world stock markets").

2.4.3 Detrended Price Oscillator (DPO):

The Detrended Price Oscillator (DPO) is a technical indicator that helps us focus on natural ups and downs in stock prices by removing the overall long-term trend isolate and highlighting shorter-term price cycles and fluctuations. By calculating the difference between the current price and a displaced moving average to backward, the DPO allows traders to identify recurring patterns, estimate the duration of price cycles between peaks and troughs, and spot overbought or oversold conditions within those cycles. In practical terms, the DPO lets us see when prices have reached a typical peak or bottom, helping us guess when a change in direction might occur. This is useful for timing trades and understanding when the market might reverse or continue its pattern (Grafton).

2.4.4 Future Line of Demarcation (FLD):

The Future Line of Demarcation (FLD) is a cycle analysis indicator developed by Hurst. By moving the price series forward by the length of the dominant cycle, we projected the continuation of the cycle into the future. The intersection between price and FLD lines identifies potential reversal zones (crests and troughs). This technique provides a predictive visualization of turning points and is widely used in cycle-based trading systems to anticipate inflection dates and optimize entry/exit timing (Grafton); (Alsalloum).

This triple-layered approach, FFT for frequency discovery, DPO for cycle isolation, and FLD for trajectory projection, offers a comprehensive view of multiscale cyclical behavior within the TASI index.

2.4.5 Hurst Exponent (H)

The Hurst exponent *(H)* is a statistical measure that tells us how much today's price changes in the market are connected to past price movements that used to characterize the long-term memory or persistence of a time series, offering a statistical lens to evaluate market efficiency and price persistence. This study integrates both static and dynamic Hurst exponent computations to characterize and monitor the evolving regime behavior of the TASI equity index.

Rescaled range (R/S) analysis is a way to measure how much a series of data moves away from its average over different time periods. It works by taking the difference between the highest and lowest values of the 'cumulative sum' of how each value differs from the average and then dividing it by the usual amount the data typically vary (the standard deviation). This helps us to see whether a market tends to continue moving in the same direction, reverse course, or act randomly over time.

The rescaled range *(R/S)* for a time series $X_1, X_2, \ldots, X_n$ is calculated as:

$$\frac{R(n)}{S(n)} = \frac{\max(Z_1, Z_2, \ldots, Z_n) - \min(Z_1, Z_2, \ldots, Z_n)}{S(n)}$$

where

$$Z_t = \sum_{i=1} (X_i - \bar{X}) \quad \text{for } t = 1, 2, \ldots, n$$

and S(n)is the standard deviation of the series. Looking at how this R/S value changes for longer or shorter periods and plotting the results, we can estimate the Hurst exponent, which tells us about long-term memory or trends in the data. A high value means that past trends often continue; a low value means that the series tends to reverse itself quickly (Hurst); (Peters); (Zournatzidou and Floros).

Calculated using the classical **Rescaled Range (R/S)** analysis on log-return series, static H values provide baseline insights into market regimes:

- $H > 0.5$: it means the market is showing trends or 'memory', so past moves often continue. Indicates persistent trends and momentum effects, typical of sentiment-driven bull markets or inefficient pricing.
- $H < 0.5$: prices tend to move in the opposite direction more often, or quickly return to normal levels. Suggests mean-reverting or corrective behavior, potentially exploitable by contrarian strategies or value-based investors.
- $H = 0.5$: means the market is acting totally random, like flipping a coin. Implies a pure random walk consistent with efficient market behavior.

The R/S method involves partitioning time series into intervals, computing cumulative deviations from the mean, and regressing the rescaled range against the logarithm of time intervals to estimate *H* (Di Sciorio et al.); (Cajueiro and Tabak, "The Hurst exponent over time: Testing the assertion that

emerging markets are becoming more efficient"); (Geweke and Porter-Hudak).

2.4.6 Dynamic Hurst Exponent (Rolling H_t)

To assess evolving market efficiency and adaptive behavior, a rolling 126-day window is used to compute time-varying Hurst exponents. This dynamic technique captures changes between trending and mean-reverting regimes over time, particularly valuable in the presence of regulatory interventions, macroeconomic shocks, or geopolitical uncertainty common in GCC markets (Anonymous); (Morales et al.). The rolling framework H provides early warning signals for regime changes, aligning with the fractal market hypothesis (FMH), which posits that market structures evolve in response to changing participant horizons and external stressors (Grafton); (Di Sciorio et al.). Visualization of dynamic *H* alongside price series can reveal clustering of persistence or anti-persistence, offering a novel layer of risk management and forecasting precision.

2.4.7 Rationale for Hurst Integration

Combining spectral and fractal analysis methods provides a more comprehensive view of the forces driving market cycles, particularly when sentiment and structural factors interact. Spectral analysis, such as Fast Fourier Transform (FFT), detects and quantifies repeating deterministic patterns in price data, allowing researchers to identify dominant cycles and periodicity present in financial markets (Geweke and Porter-Hudak); (Kaur and Bharucha). In parallel, fractal measures such as the Hurst exponent offer a way to capture underlying stochastic, or random, behaviors, revealing if a market is characterized by persistence (trend-following), anti-persistence (mean-reversion), or randomness over time (Vogl); (Alvarez-Ramirez et al.). By integrating these two metrics, analysts can bridge the gap between predictable rule-based cycles and probabilistic long-term memory effects, enabling a richer and more robust interpretation of how investor sentiment and market structure influence price dynamics. This hybrid approach has been shown to enhance trading signals and risk assessment for both active short-term traders and long-term strategic investors (Nazarova, Dovlabdekova, et al.); (Grafton).

3 Results

3.1 Sentiment Analysis Model Results

3.1.1 Sentiment Distribution

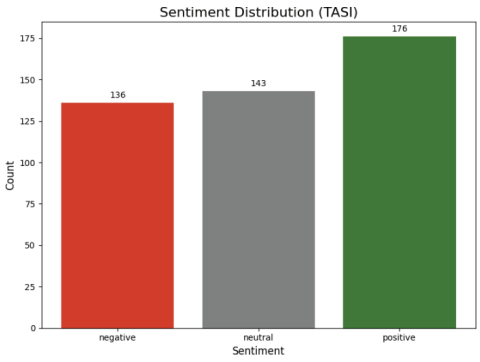
After applying the set of pre-trained Arabic sentiment analysis models (AraBERTv2, CAMeLBERT-DA, and MARBERTv2) to the collected Twitter dataset, the overall distribution of detected sentiments among users discussing the Saudi market (TASI) was quantified. The models classified each tweet into one of three categories: positive, neutral, or negative.

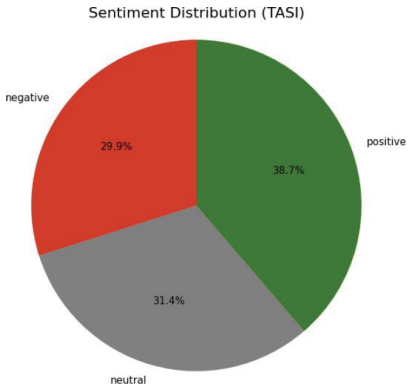
Table 3.1: Sentiment Distribution of Arabic Tweets Referencing TASI

Sentiment	Count	Percentage (%)
Positive	176	38.7
Neutral	143	31.4
Negative	136	29.9

Figure 3.1 presents both the raw counts (bar chart) and proportions (pie chart) for each sentiment class in the analyzed sample.

Figure 3.1: Distribution of sentiment labels (negative, neutral, positive) in Arabic tweets referencing TASI, as detected by the ensemble model. (a) shows Bar Chart; (b) shows Pie Chart.





3.1.2 Descriptive Statistics for TASI

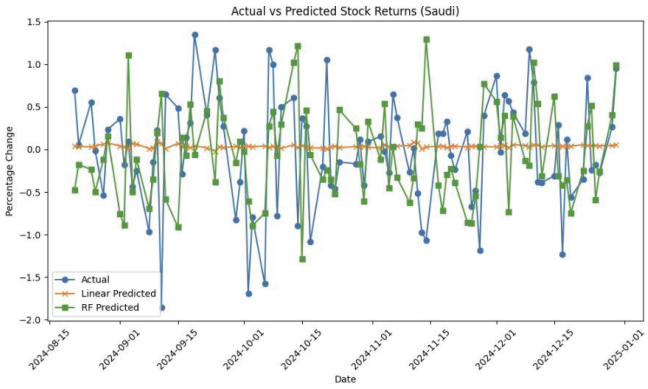
Table 3.2 presents descriptive statistics for the TASI daily closing price and percentage daily change, based on 498 trading days from January 1, 2023 to December 30, 2024.

Table 3.2: Descriptive Statistics for TASI Daily Close Price and Percentage Change (2023–2024)

Statistic / Date	Close	%Chg
Count	498	498
Start Date	2023-01-01	31.4
End Date	2024-12-30	29.9
Mean	3,084.25	0.03
Std (Standard Deviation)	175.33	0.71
Min (Lowest Value)	2,656.26	-2.42
25% (First Quartile)	2,959.64	-0.41
50% (Median)	3,117.45	0.05
75% (Third Quartile)	3,219.22	0.49
Max (Highest Value)	3,422.66	2.55

Figure 3.2 plots the actual daily percentage changes for TASI against the predicted values of both models over time.

Figure 3.2: Actual vs Predicted Daily Percentage Changes for TASI



3.1.3 Regression Model performance for TASI

Table 3.3 presents the performance metrics for Linear Regression and Random Forest models applied to daily TASI returns. Two metrics are reported: The mean absolute error (MAE), representing the average magnitude of the prediction errors, and the root mean squared error (RMSE), which gives additional weight to larger errors. Table 3.4 presents a sample of daily TASI returns, side by side with predictions made by Linear Regression and Random Forest models, including the computed residuals (difference between actual returns and model predictions) for each.

Table 3.3: Comparison of Regression Model Performance for TASI Returns

Model	MAE	RMSE
Linear Regression	0.518	0.657
Random Forest	0.586	0.790

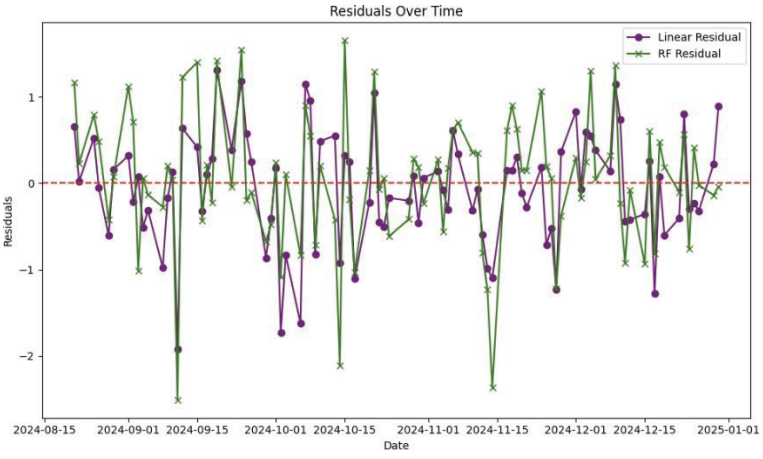
Table 3.4: Sample of Actual vs Predicted Daily Percentage Changes for TASI

Date	Actual	Linear Predicted	RF Predicted	Residual (Lin)	Residual (RF)
2024-08-21	0.696	0.038	-0.473	0.658	1.170
2024-08-22	0.057	0.035	-0.177	0.022	0.234
2024-08-25	0.557	0.030	-0.236	0.526	0.793
2024-08-26	-0.014	0.039	-0.498	-0.054	0.484
2024-08-28	-0.542	0.065	-0.115	-0.607	-0.426

3.1.4 Residuals Over Time

Figure 3.3 presents the time series of residuals (prediction errors) for both the Linear Regression (shown as purple circles and lines) and Random Forest (green crosses and lines) models over several months of TASI data.

Figure 3.3: Residuals Over Time: Difference between actual and predicted TASI returns for Linear Regression (purple) and Random Forest (green) models. The red dashed line at zero highlights unbiased prediction.



Residuals represent the difference between the actual market return and what the model predicted. When the residuals are close to zero, the predictions are accurate; larger numbers mean that the model was missed by a larger amount (Verma). Both the Random Forest and Linear Regression models show residuals that bounce above and below zero, but without any lasting pattern—meaning that neither consistently over- or underestimates for long periods. This indicates that both models generally avoid systematic bias, which is a good sign of reliability (Kalusivalingam et al.). However, the Random Forest model produces a wider spread of errors, occasionally making larger mistakes, while the Linear Regression model keeps most

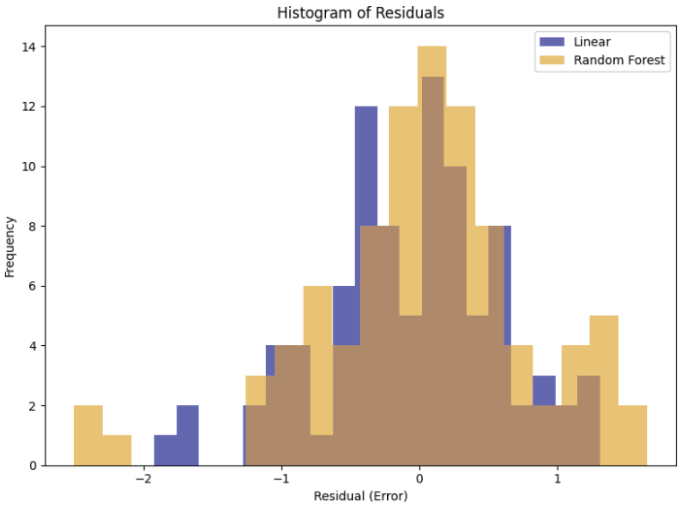
of its errors within a smaller range. This shows that Random Forest is more sensitive to short-term ups and downs, so its predictions can be more volatile compared to steady predictions from Linear Regression (Mendez and Lohr); (Jonsson and Fredrikson); (Jui et al.).

Both models experience periods of high volatility in their prediction errors, particularly during times of market turbulence, such as September–October and mid-November. These clusters likely correspond to the major news or events in which the prediction of the outcomes becomes much harder (Jain et al.); (DeRousseau et al.). The Linear Regression model generally produces errors that are small and close to zero, in line with its tendency to make stable, average-focused predictions. In contrast, the Random Forest model more readily reacts to large swings, which sometimes leads to larger mistakes (Jonsson and Fredrikson); (Limbong et al.); (Jain et al.). Despite these differences, both models show residual errors that balance above and below zero over time, suggesting that neither systematically over- or underestimates for long stretches, a sign of no systematic bias. This analysis highlights that, while Linear Regression is more consistent, it may miss extreme moves, whereas Random Forest is more flexible but riskier in turbulent periods. These residual patterns signal that predicting sudden market shocks remains very difficult and that combining new information or more advanced models could improve future forecasts (Jain et al.); (DeRousseau et al.); (Jonsson and Fredrikson).

3.1.5 Histogram of Residuals

Figure 3.4 displays the distribution of residuals (prediction errors) for both the Linear Regression (blue) and Random Forest (orange) models. Each residual is calculated as the difference between the actual TASI return and its predicted value for that day.

Figure 3.4: Histogram of residuals (prediction errors) from Linear Regression and Random Forest models for TASI returns.



3.1.6 Classifier Performance

Table 3.5: Performance Comparison of Sentiment Classification Models on TASI Arabic Tweets

Classifier	Metric	Count Occurrence	TF-IDF
	Accuracy	0.669	0.673
Logistic Regression	F1-Weighted	0.669	0.674
	F1-Macro	0.670	0.674
	F1-Micro	0.669	0.673
	Accuracy	0.634	0.646
Naive Bayes	F1-Weighted	0.634	0.645
	F1-Macro	0.634	0.644
	F1-Micro	0.634	0.646

Table 3.5 shows how well two different models, logistic regression and naive Bayes, can tell if a tweet is positive, negative, or neutral. To do this, we used two ways to turn tweets into numbers the models can read: by counting how often words appear (count events) and by using a method called TF-IDF that also weighs rare words more heavily. We checked each model's performance using four common metrics: accuracy (how often it gets the answer right), F1 weighted (balances accuracy across all categories), F1-macro (average performance for each category, treating all classes equally), and F1-micro (overall performance, giving more weight to common classes). These measures help us to understand not just whether the models are right, but also how well they handle different types of sentiment in the data (Farha and Magdy); (Huang and Li); (He et al.); (Zhou et al.).

3.1.7 Model Tuning and Evaluation

To build the sentiment analysis model, we tried eight different sets of settings (called hyper-parameters) for the linear SVM classifier to see which combination worked best. Each time, the model was trained and checked using three data splits, a process known as cross-validation, making sure that our results were reliable and not just lucky for one group. The settings we tested included the C value, which controls how flexible the model is and how we balance the different sentiment classes. The most accurate and fair results came from the set model at C = 0.5 and the balanced weights, helping it to correctly handle all types of sentiment without favoring one over another (Farha and Magdy); (Sakınç et al.);(Alosaimi et al.). These findings are consistent with best practices for using SVM in text and sentiment classification in research and industry (Iddrisu et al.); (Han et al.).

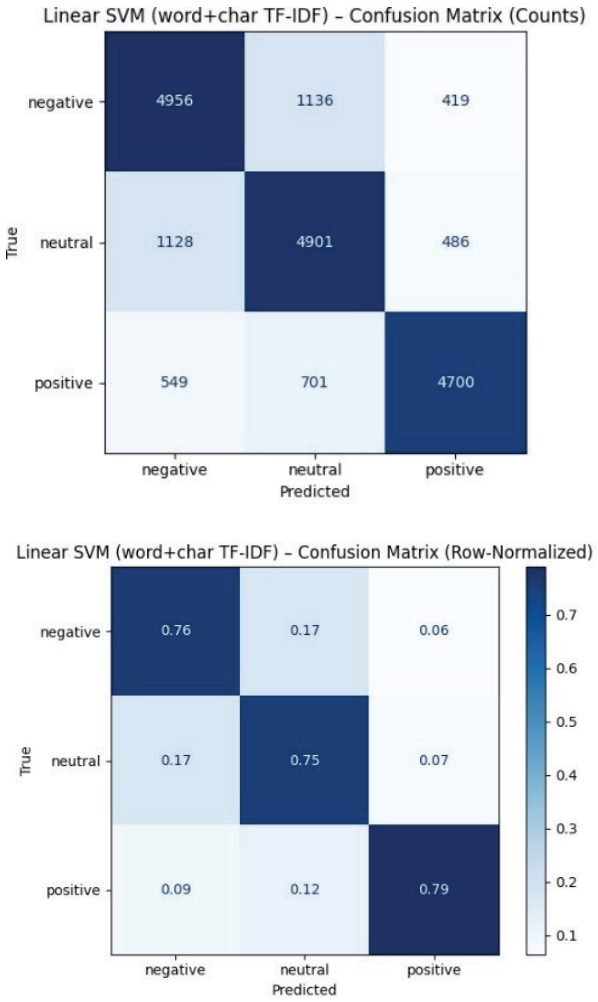
Table 3.6: Performance of the Tuned Linear SVM Sentiment Model

Class	Precision	Recall	F1-score	Support
Negative	0.747	0.761	0.754	6,511
Neutral	0.727	0.752	0.740	6,515
Positive	0.839	0.790	0.814	5,950
Accuracy			0.768	18,976
Macro avg	0.771	0.768	0.769	18,976
Weighted avg	0.769	0.767	0.768	18,976

Table 3.6 reports precision, recall, and F1-score for each sentiment class and aggregated averages for the chosen SVM model.

Figures 3.5 show the results counts and the normalized percentages in rows for each pair of true / predicted labels.

Figure 3.5: Confusion matrix for the final Linear SVM model: above, raw counts; below, row-normalized percentages.



3.1.8 Confusion Matrix and Model Comparison

Table 3.7: Confusion Matrix (Counts) for Sentiment Classification

	Predicted Negative	Predicted Neutral	Predicted Positive
True Negative	4,956	1,136	419
True Neutral	1,128	4,901	486
True Positive	549	701	4,700

After choosing the best model, we found that using a linear SVM with word- and character-based TF-IDF features was the most effective in classifying the sentiment of the tweet. This approach helps the computer understand not just the main words, but also the smaller patterns within the text, so it can better decide if a tweet is positive, negative, or neutral. The model results are shown in tables and charts that clearly show how well it performed. Our findings are consistent with recent studies showing that the combination of SVMs with TF-IDF characteristics leads to strong results for the analysis of Arabic and financial sentiment (Farha and Magdy);(Sakınç et al.);(Alosaimi et al.); (Duet al.);(Iddrisu et al.); (Han et al.).

Table 3.8: Confusion Matrix (Normalized by True Label)

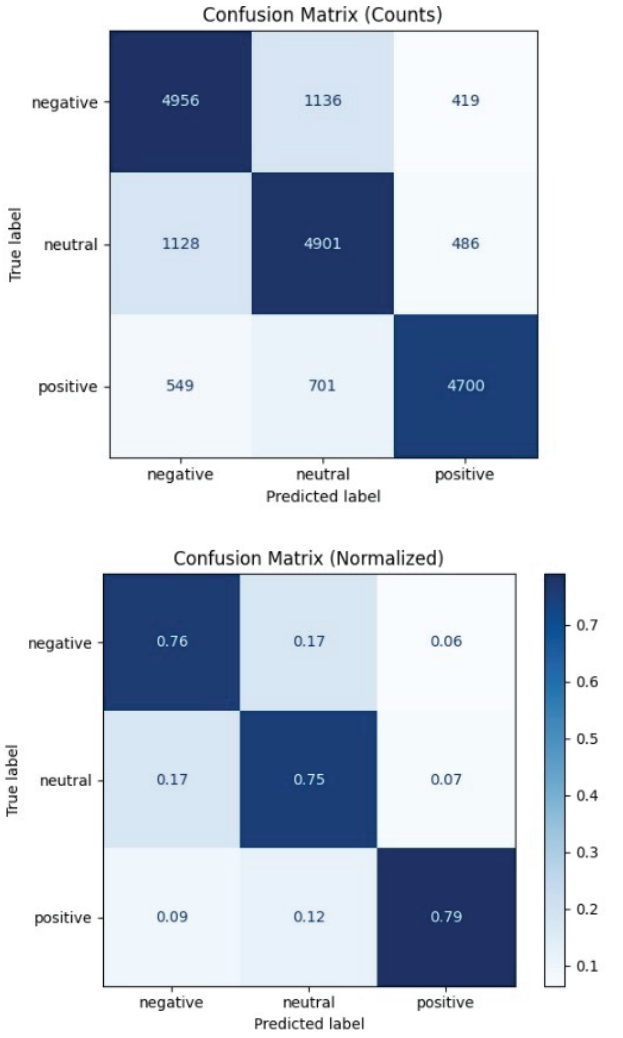
	Predicted Negative	Predicted Neutral	Predicted Positive
True Negative	0.76	0.17	0.06
True Neutral	0.17	0.75	0.07
True Positive	0.09	0.12	0.79

Table 3.9: Classification Report: Sentiment Analysis Model Performance (SVM)

Class	Precision	Recall	F1-score	Support
Negative	0.747	0.761	0.754	6,511
Neutral	0.727	0.752	0.740	6,515
Positive	0.839	0.790	0.814	5,950
Accuracy			0.768	18,976
Macro avg	0.771	0.768	0.769	18,976
Weighted avg	0.769	0.767	0.768	18,976

Tables 3.7 and 3.8, with Figure 3.6, show the distribution and precision of the sentiment predictions for the negative, neutral, and positive classes. Table 3.9 reports the precision, recall, and F1-score for each class, as well as the general and average results.

Figure 3.6: Confusion matrices for sentiment classification. Top: Raw number of tweets in each class (negative, neutral, positive). Bottom: Normalized confusion matrix (row percentages). Both show how well the model classified each sentiment.



3.1.9 Model Uncertainty and Common Misclassifications

To find the parts of our sentiment model that worked less well, we looked at two things: predictions where the model was least sure (those with the lowest confidence scores), and cases where it often mixed up one type of sentiment for another, such as confusing positive tweets with neutral ones. Analyzing these tough cases helps us spot patterns and understand where the model needs more improvement. Research shows that checking low-confidence predictions and common mistakes is a key step in making sentiment models more reliable and useful in finance and social media studies (Farha and Magdy); (Sakınç et al.); (Alosaimi et al.); (Mehta and Pandya); (Huang and Li).

Table 3.10: Misclassifications Analysis: Least Confident vs Most Frequent Results

Tweet ID	True	Pred.	Margin Gap		True → Pred.	Count
16,348	pos	neu	0.000045		neg → neg	4956
12,392	pos	neu	0.000222		neu →neu	4901
1,867	pos	neu	0.000268		pos →pos	4700
5,728	neg	neu	0.000282		neg →neu	1136
16,391	neu	neg	0.000333		neu →neg	1128
13,546	pos	neg	0.000432		pos →neu	701
1,373	neg	pos	0.000859		pos → neg	549
11,731	neg	neu	0.000906		neu →pos	486
13,170	pos	neu	0.000965		neg →pos	419
15,332	neg	pos	0.001089			

Table 3.10 summarizes the lowest confidence predictions (smallest margin gaps) and the most frequent correct and incorrect predictions among classes in the test set.

3.1.10 Performance of Logistic Regression with CountVectorizer

Table 3.11 shows how well the Logistic Regression model was able to detect positive, negative, and neutral sentiments in Arabic tweets about TASI. For each type of mood, we measured precision (how often the model's positive or negative label was correct), recall (how often were all positive, negative, or neutral tweets found) and the F1 score (which combines precision and recall for a balanced rating). The model used a bag-of-words approach, which counts how often words appear in each tweet to help make its predictions. The table gives both numbers for each mood and overall averages, making it easy to see where the model did well and where it needs improvement. Studies confirm that this kind of setup, using logistic regression on word counts, is common in Arabic sentiment analysis and provides reliable results in real-world applications (Farha and Magdy); (Alosaimi et al.); (Al-Bataineh); (Albahli); (Aladeemy et al.).

Table 3.11: Classification Report for Logistic Regression (CountVectorizer Features)

Class	Precision	Recall	F1-score	Support
Negative	0.76	0.78	0.77	13,251
Neutral	0.57	0.62	0.59	6,705
Positive	0.72	0.64	0.68	7,435
Accuracy			0.70	27,391
Macro avg	0.68	0.68	0.68	27,391
Weighted avg	0.70	0.70	0.70	27,391

3.1.11 Stratified K-Fold Cross-Validation

To ensure that the model treated each kind of tweet fairly, we used something called Stratified K-Fold Cross-Validation. This means that we split our data into several groups ("folds"), so each group kept the same mix of positive, neutral, and negative tweets as the original set. By doing this, when we trained and tested the model, we could be sure that it was learning and predicting each type of sentiment in a balanced way, not favoring any one class. Research shows that using stratified cross-validation is important for reliable results in sentiment analysis and helps prevent bias when dealing with different classes in social media data (Farha and Magdy);(Alosaimi et al.).

Table 3.12 summarizes the averages (means) and variability (standard deviations) for each performance score on all validation folds.

Table 3.12: Stratified K-Fold Cross-Validation Results (Mean ± Standard Deviation)

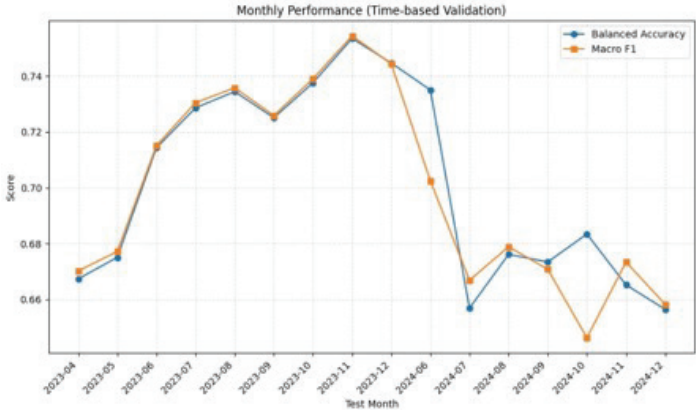
Metric	Mean	Std. Dev.
Balanced Accuracy (bal_acc)	0.7613	0.0033
F1 Macro (f1_macro)	0.7627	0.0034
F1 Weighted (f1_weighted)	0.7614	0.0033
Precision Macro (precision_macro)	0.7648	0.0035
Recall Macro (recall_macro)	0.7613	0.0033
Accuracy (accuracy)	0.7608	0.0033

3.1.12 Time-based Validation

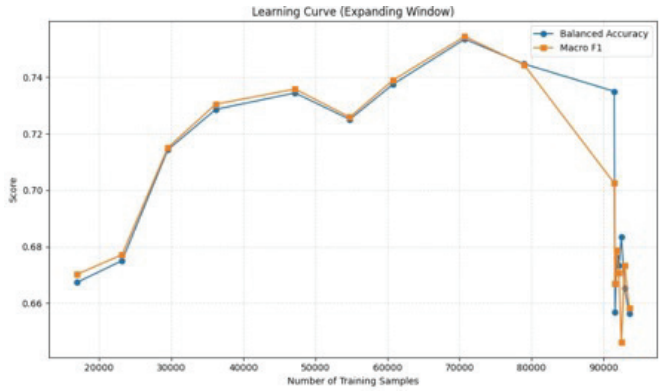
To make sure our model could keep up with changes in social media over time, we used a special method called time-based, or walk-forward, validation. This means that the model was trained using all the tweets up to each month, then it was tested on the tweets from the next month, doing this step by step, month after month, starting from April 2023. This lets us see how well the model adapts to new trends and fresh data in real time. Studies confirm that walk-forward validation is especially useful for tracking and predict-ing evolving patterns in time-based social networks and financial data (Mahesh et al.); (Beniwal et al.); (Latif et al.).

Table 3.13 summarizes monthly follow-up test scores, as training data increases and predictions are made for unseen months.

Figure 3.7: Model evaluation plots: (a) Monthly performance across validation splits, (b) learning curve showing impact of training size.



(a) Monthly Performance (Time-based Validation). Blue: balanced accuracy; orange: macro F1 for each test month.



(b) Learning Curve (Expanding Window). Both scores rise as more training data accumulates.

Table 3.13: Walk-Forward (Time-Based) Validation Results: First 10 Monthly Splits

Train Period	Test Month	Train Size	Test Size	Balanced Acc.	F1 Macro	F1 Weighted	Accuracy
2023-01 → 03	2023-04	16,887	6,297	0.67	0.67	0.67	0.67
2023-01 → 04	2023-05	23,184	6,354	0.68	0.68	0.68	0.68
2023-01 → 05	2023-06	29,538	6,648	0.71	0.72	0.72	0.71
2023-01 → 06	2023-07	36,186	10,998	0.73	0.73	0.73	0.73
2023-01 → 07	2023-08	47,184	7,611	0.73	0.74	0.74	0.73
2023-01 → 08	2023-09	54,795	6,033	0.73	0.73	0.73	0.73
2023-01 → 09	2023-10	60,828	9,907	0.74	0.74	0.74	0.74
2023-01 → 10	2023-11	70,735	8,231	0.75	0.75	0.75	0.75
2023-01 → 11	2023-12	78,966	12,358	0.74	0.74	0.75	0.75
2023-01 → 12	2024-01	91,324	12	0.50	0.48	0.62	0.67

3.2 Market Cycle Analysis Results

3.2.1 FFT Spectral results: Determining Market Cycle Periods

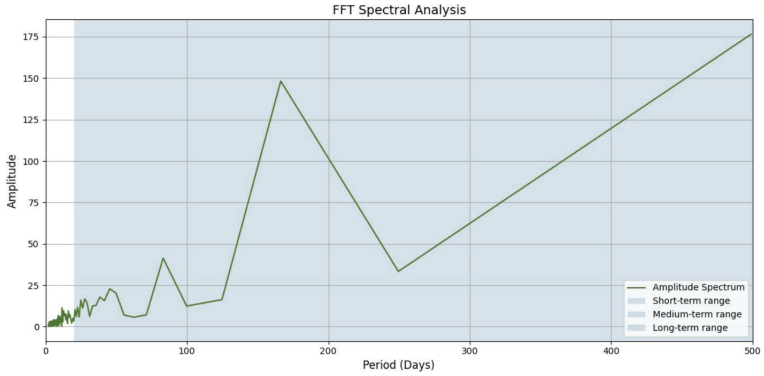
To find repeating rhythms or cycles in the TASI market behavior, we use a method called Fast Fourier Transform (FFT) on its price data over time. FFT is a mathematical tool that helps us break down complicated signals, such as daily changes in stock prices, into simple waves of different speeds. By doing this, it becomes easier to spot regular patterns in the market, such as cycles or trends that repeat over weeks or months. This technique is widely used in financial research to uncover hidden market cycles and has been shown to be effective in analyzing data from the stock market (Kaur and Bharucha); (Cajueiro and Tabak, "Testing for long-range dependence in world stock markets"); (Alosaimi et al.); (Refinitiv).

Table 3.14: Detected Market Cycles from FFT Analysis

Cycle Type	Cycle Period (Days)
Short-term cycle	45
Medium-term cycle	166
Long-term cycle	499

Table 3.14 summarizes the main cycle period detected. Figure 3.8 shows the resulting amplitude spectrum, where the x-axis (period) indicates possible cycle lengths (days), and the y-axis (amplitude) indicates the strength of each periodicity.

Figure 3.8: FFT amplitude spectrum for TASI market data. Peaks indicate dominant cycle lengths. Shaded background highlights short-, medium-, and long-term cycle ranges.



3.2.2 Detrended Price Oscillator (DPO) and FLD Forecasting: Cycle Confirmation and Turning Point Projections

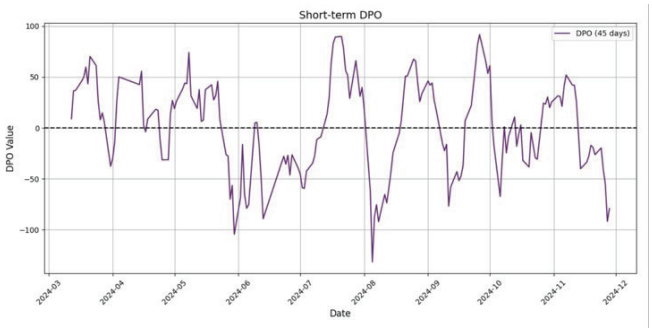
After finding the main cycle lengths with FFT, we used the Detrended Price Oscillator (DPO) to make these rhythms easier to see on a chart by removing longer-term price trends and highlighting the repeating ups and downs. This tool allows us to spot regular peaks and valleys in TASI prices by focusing only on short-term cycles instead of general movement (Lang et al.); (Q. Chen et al.); (Farha and Magdy). Next, we applied the Future Line of Demarcation (FLD) method, which uses these cycles to estimate when the next peak or drop might occur and what price it could reach. Combining DPO and FLD helps us forecast future market turning points more clearly, giving investors a practical look at possible price movements in advance (Refinitiv); (Cajueiro and Tabak, "Testing for long-range dependence in world stock markets"); (Alosaimi et al.).

The following figures show, for each cycle, the corresponding DPO (left) and FLD (right) forecasts:

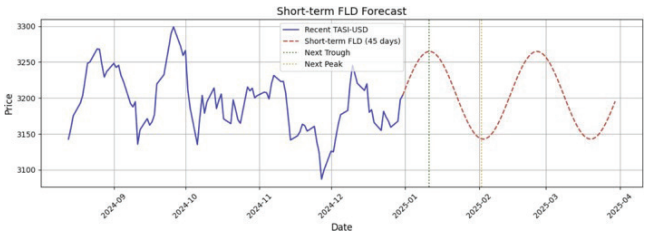
Table 3.15: Cycle Turning Points Detected and Forecast with DPO and FLD

Cycles	Period	Last Trough	Next Trough	Next Peak	Target (Price)
Short-term	45d	2024-11-27	2025-01-11	2025-02-02	3142.60 (trough), 3264.90 (peak)
Medium-term	166d	2024-08-05	2025-01-18	2025-04-11	3028.16 (trough), 3379.34 (peak)
Long-term	499d	2023-12-31	2026-09-24	2027-05-31	Out of period

Figure 3.9: Short-term cycle analysis using DPO and FLD.

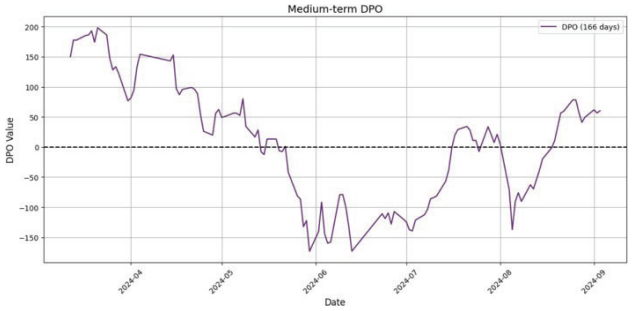


(a) Short-term (45-day) DPO.

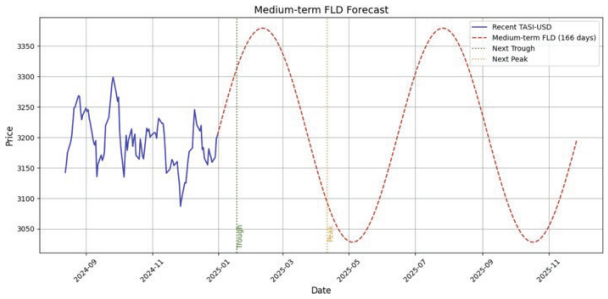


(b) Short-term FLD forecast.

Figure 3.10: Medium-term cycle analysis using DPO and FLD.



(a) Medium-term (166-day) DPO.

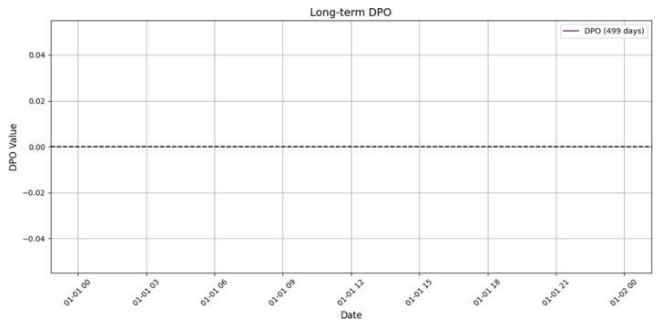


(b) Medium-term FLD forecast.

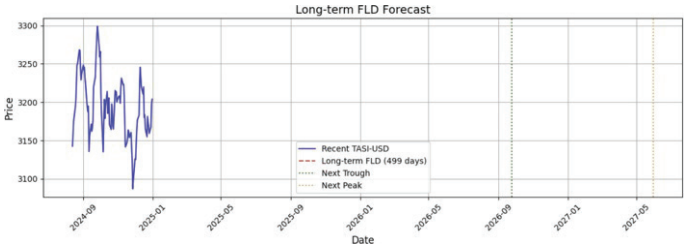
3.2.3 Hurst Exponent

To ensure that our method for finding cycles and forecasting price changes in TASI was reliable, we checked for something called a 'confluence check'. In this context, confluence means that the different cycles we found, using independent tools, predicted the same major turning point in time, like a price peak or bottom. When several cycles align in this way, it adds extra confidence that our forecast is correct. Financial studies show that looking for confluence, or agreement between multiple signals, is a trusted way to strengthen predictions and reduce errors in market analysis (Farha and Magdy); (Refinitiv); (Uddin).

Figure 3.11: Long-term cycle analysis using DPO and FLD.



(a) Long-term (499-day) DPO.



(b) Long-term FLD forecast.

Table 3.16: Upcoming Cycle Turning Points and Confluence Check

Cycle	Type	Next Trough	Next Peak
Short-term	45 days	2025-01-11	2025-02-02
Medium-term	166 days	2025-01-18	2025-04-11
Long-term	499 days	2026-09-24	2027-05-31

Table 3.16 lists the next fore-casted trough and peak of the short-, medium-, and long-term cycles. Both short-term and medium-term cycles predict price troughs that occur within 7 days of each other in January 2025.

3.2.4 Fractal Market Hypothesis Results: Rolling Hurst, Volatility, and Stability

The Fractal Market Hypothesis (FMH) is a theory that suggests financial markets behave in different ways over different periods of time because investors have varying strategies and goals. Some focus on quick trades, while others think about long-term trades (Peters); (G. Liu et al.); (Cajueiro and Tabak, "Testing for long-range dependence in world stock markets"); (Alosaimi et al.). According to FMH, during stable times, this mix of short- and long-term investors maintains the market balanced. But when there is uncertainty or a crisis, all investors can start acting in the short term, leading to big price swings and instability. This idea helps explain why markets do not always act in a predictable way and why sudden shocks can cause dramatic changes, as seen in TASI and other global markets (Sakınç et al.); (Uddin).

Table 3.17: Cycle Stability Metrics for FMH Assessment

Cycle	Length (Days)	Amplitude Std	Amplitude Mean	Stability Ratio
Short-term	45	40.77	34.42	1.18
Medium-term	166	117.06	99.64	1.17
Long-term	499	-	106.98	-

Figure 3.12: Rolling Hurst exponent for the TASI Index (126-day window). The dashed line at $H = 0.5$ represents a random walk benchmark. Values above this line indicate periods of trend persistence.

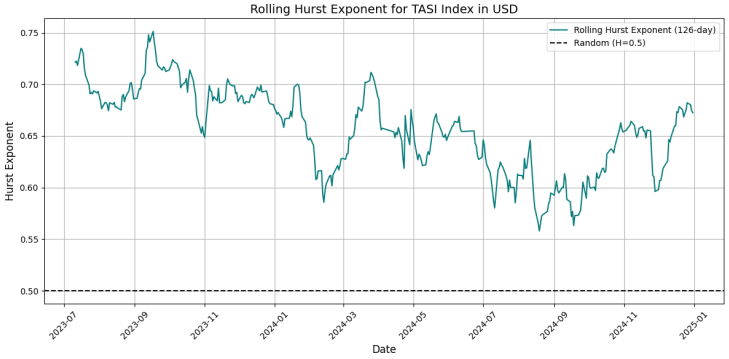


Table 3.17 shows the amplitude, mean and stability ratio (std/ mean) for each detected cycle. Figure 3.12 displays the rolling Hurst exponent. Figure 3.13 presents the short-term volatility period-by-period for the TASI index.

Figure 3.13: Short-term (20-day) annualized volatility for the TASI Index. Spikes indicate episodes of heightened instability.



Chapter 4 Discussion

4.1 Sentiment Analysis Model Interpretation

4.1.1 SEntiment Distribution

As shown in Table 3.1, the results of the sentiment classification for Arabic tweets about the Saudi market (TASI) reveal that positive sentiment was the most common, accounting for 38. 7% of all tweets. This suggests that during the study period, people online felt fairly optimistic or 'bullish' about the market. Neutral sentiment was also high (31.

4%), indicating that much of the discussion was balanced or undecided, while negative sentiment was not far behind at 29.9%, showing that critical and pessimistic opinions were still a significant part of the conversation.

Figure 3.1 helps visualize these findings by presenting both a bar chart and a pie chart. The bar graph shows the number of tweets in each mood and the pie chart illustrates the share of each sentiment in the total. Together, these visuals confirm that while positive perspectives were a little more frequent, negative and neutral views were also strongly represented. This relatively uniform split between the three types of sentiment highlights the diversity and dynamic nature of investor opinions expressed in Saudi financial social networks.

Such sentiment distributions, as seen in Table 3.1 and Figure 3.1, are valuable for understanding how confident or concerned investors feel, as well as potential crowd behavior and market psychology. Recent studies on financial sentiment analysis confirm that tracking these proportions can help explain market trends and support investors in making better decisions.

4.1.2 TASI Descriptive Statistics

Table 3.2 gives a clear summary of how the TASI index performed on 498 trading days from January 2023 to December 2024. The average closing price in this period was 3,084.25, which means that TASI generally finished each trading day around this level. Prices ranged over time, reaching a low of 2,656.26 and reaching a high of 3,422.66, reflecting not only big-picture trends, but also sudden market changes.

The standard deviation of closing prices (175.33) listed in the table measures the variability or 'volatility' in daily price levels. Higher standard deviations, such as this, are typical for active or uncertain markets and signal that price movements, sometimes small, sometimes very large, were common. These swings can result from major economic news, changes in investor confidence, or global events, as highlighted in recent research on the prediction of TASI volatility.

The mean daily percentage change was only 0. 03%, which means that the index did not move steadily up or down during the study period; instead, it hovered near the same level most days. The standard deviation for daily changes (0.71%), and observed minimum and maximum movements of -2.42% and +2.55%, also shown in the table, show that while most days had small changes, some days saw large jumps or drops.

The quartile values in the table further highlight market stability: Half of all closing prices were between about 2,960 and 3,219, while daily returns for most days fell between -0.41% and +0.49%. This pattern, a mostly stable market interrupted by occasional strong moves, is typical for emerging or reforming markets and is consistent with recent studies on TASI and similar financial indices.

In summary, Table 3.2 shows that TASI was generally steady during the study period, but sometimes experienced sharp changes, giving a balanced picture of both stability and volatility in the Saudi market.

4.1.3 Regression Model Performance

As shown in Table 3.3, the Linear Regression model made more accurate predictions than the Random Forest for daily TASI returns. Its mean absolute error (MAE) was 0.518 and the root mean square error (RMSE) was 0.657, both lower than the Random Forest scores of 0.586 (MAE) and 0.790 (RMSE). This means that Linear Regression generally stayed closer to the real market changes and made smaller mistakes on average. Although Random Forest is able to handle more complex relationships, it sometimes gets distracted by short-term noise

in the data, causing bigger errors.

Table 3.4 gives specific examples, showing the actual market changes next to what each model predicted, along with their respective errors (residuals). You can see that on days with bigger actual market moves (like 2024-08-21 and 2024-08-25), both models struggled, but Random Forest often made larger mistakes than Linear Regression.

Figure 3.2 visualizes the predictions and the actual market values over time, making it clear that both models have difficulty tracking the biggest swings. Linear regression predictions are more stable and stick closer to the average, while Random Forest sometimes follows the direction of large moves, but does not always match their size.

In summary, as the tables and figures show, linear regression is a more reliable and consistent starting point for linking sentiment signals to TASI price movements, especially when the data are relatively smooth. Complex models like Random Forest can sometimes "chase the noise", reacting to patterns that do not repeat and resulting in less reliable predictions. This pattern is common in financial research and supports the use of simpler methods when the goal is clear and stable forecasting.

4.1.4 Actual vs Predicted Daily Change Analysis

Examining Table 3.4, we see that both the Linear Regression and Random Forest models struggle when the market experiences sudden big jumps or drops in daily returns. Linear regression tends to make stable predictions that stay close to zero, even on days like 2024-08-21 and 2024-08-25 when the actual market moved sharply. This leads to large errors during these volatile periods. Random Forest sometimes catches the direction of these big changes, but often underestimates how big they really are, which results in its own large mistakes.

Figure 3.2 makes these challenges clear: the blue line showing actual TASI returns has many fast and unpredictable 'spikes' that neither model can track well. Linear regression predictions are flatter, missing dramatic moves, while Random Forest predictions react more but still do not match the size of real market swings.

Figure 3.3 provides a detailed look at the errors (called residuals) for both models over several months. The purple line shows Linear Regression errors, which cluster closer to zero for most dates; the green line shows Random Forest errors, which swing wider and more often, reflecting its greater volatility. Both models have errors above and below the red zero line, which means that they do not systematically over- or under-predict over time, a sign that they are unbiased, as highlighted in recent forecasting studies. In particular, during high volatility periods like September—October and mid-November, both models make bigger mistakes, likely because of difficult-to-predict market events.

Overall, the residual plots and sample prediction table show that Linear Regression consistently makes smaller, steadier errors, while Random Forest is more sensitive to quick market changes but can make bigger mistakes, especially in turbulent times. These findings, supported by other research, highlight that sudden market shocks are hard to predict and that stronger results may come from adding more market information or using advanced models like deep learning in future studies.

4.1.5 Residual Histogram

Figure 3.4 shows the distribution of prediction errors (residuals) from the Linear Regression (blue) and Random Forest (orange) models for the TASI returns. Most residuals for both models are grouped close to zero on the chart, which means that the majority of predictions were fairly accurate. Both models make mistakes above and below zero with no consistent bias, so they do not habitually over- or underestimate real results.

Looking closer at the histogram, the linear regression errors form a taller and narrower shape, showing that its predictions usually miss only a little. This "conservative" pattern means that linear regression errs carefully and rarely makes large mistakes. In contrast, Random Forest's errors are spread more widely and flatter, which means that it sometimes reacts to big short-term changes and can make larger and more frequent mistakes.

So, for day-to-day forecasting, Linear Regression is a steadier and more reliable choice, especially when the market is stable. Random Forest is more flexible for spotting unusual patterns, but, as the histogram shows, this can lead to bigger errors if the market is noisy or unpredictable. This result matches what is found in other studies of financial prediction: simpler models like Linear Regression are good at keeping errors under control, while more complex models trade some of that safety for greater adaptability.

For sentiment classification, Table 3.5 shows the results for logistic regression and naive Bayes when analyzing Arabic tweets about TASI. Logistic regression performed consistently better across all metrics, with scores for accuracy and all three measures of F1 higher than Naive Bayes, with both methods of turning tweets into numbers (word count and TF-IDF). Evaluating accuracy and multiple F1 scores in the table helps ensure that the models are not only correct overall, but can also handle different types of sentiment fairly.

4.1.6 Classifier Performance

As shown in Table 3.5, Logistic Regression outperformed Naive Bayes in all categories, whether we counted words or used the more advanced TF-IDF features for sentiment analysis of Arabic tweets about TASI. For instance, using TF-IDF, Logistic Regression achieved an accuracy of 0.673 and an F1-macro score of 0.674, which are both noticeably higher than Naive Bayes' best accuracy of 0.646. This means that logistic regression was better at correctly detecting whether tweets were positive, negative, or neutral, and gave more balanced results across all moods, as the F1 scores in the table also show.

Both models performed better with TF-IDF than with simple word counts. This is because TF-IDF gives extra importance to rare or unique words that often reveal the real feeling behind a tweet - a key advantage in finance tweets, where specific terms can strongly indicate sentiment. The table also shows that accuracy, F1 micro and F1 weighted values are quite close for each approach, meaning the models did not just guess the most common mood; they were fair and evenly balanced between the different sentiment classes.

In general, Table 3.5 makes it clear that both the type of model and the way tweets are represented have a great impact on how well the system works. Logistic regression with TF-IDF was the

best combination in this comparison, but since all accuracies are below 68%, there is still much room for improvement. Advanced methods, such as deep learning or large language models, may be needed in the future to handle the more complex language and subtle clues found in Arabic finance tweets, as highlighted in recent research.

4.1.7 Model Tuning and Evaluation

Table 3.6 summarizes how well the tuned Linear SVM model classified the sentiment of Arabic tweets about TASI. The model achieved a balanced accuracy of 0.768, which means that it correctly predicted the mood of about 77% of the tweets in the test set.

As the table shows, the model was especially good at detecting positive tweets, with a precision of 0.839 and a recall of 0.790 for the positive class, the best among all types of sentiment. For negative and neutral tweets, the results were also strong, with F1-scores above 0.74 for all classes. The macro and weighted F1-averages are nearly identical to the overall accuracy (all close to 0.77), which shows that the model treated all classes quite fairly and did not favor just the most common mood.

Figure 3.5 visualizes the confusion matrices for the model. The left panel (also shown in Table 3.7) represents the raw number of correct and incorrect predictions. The right panel shows these results as percentages, which makes it easier to see patterns. Most predictions line up along the diagonal, meaning the model labels tweets correctly most of the time. When errors did occur, they were usually between neutral and one of the other classes (negative or positive). Very few tweets were directly confused between positive and negative, which is reassuring because these are the most different moods.

Together, these tables and figures highlight that the SVM model is robust: it handles the informal and sometimes confusing language of social media well, producing reliable and balanced results. This makes the model a solid tool for measuring public mood and can help future research connect collective sentiment to real changes in the Saudi market.

4.1.8 Confusion Matrix and Model Comparison

Tables 3.7 and 3.8, along with Figure 3.6, show that the linear SVM model did very well in classifying tweets as negative, neutral, or positive. In the normalized confusion matrix (Table 3.8 and Fig. 3.6 right), we see that the model correctly recognized 76% of negative tweets, 75% of neutral tweets, and 79% of positive tweets, a strong performance along the diagonal of the matrix. Only a small number '6–7%' were confused for opposite classes, and most errors occurred between neutral and either negative or positive, which is common in financial social networks due to the subtle way people express mixed feelings or uncertainty.

Table 3.9 provides a detailed breakdown of model performance, showing that precision, recall, and F1 scores are highest for positive tweets. For example, the precision for positive tweets is 0.839 and recall is 0.790, giving an F1 score of 0.814. All classes score above 0.74 for F1, and macro/weighted averages are nearly identical to overall accuracy (0.768), confirming that the model treats all sentiment types fairly and consistently.

The raw count confusion matrix (Table 3.7 and Fig. 3.6 left) further confirms that the SVM model makes mostly correct decisions, with much fewer errors in directly swapping positive and negative tweets - a common mistake with simpler models.

Together, these tables and figures clearly show that our Linear SVM—using both word and character TF-IDF features—is highly effective for sentiment analysis of Arabic tweets in the Saudi market. Strong, balanced performance, high interpret ability, and low confusion make this approach a robust tool for tracking and quantifying investor mood online, especially compared to basic models such as Naive Bayes. Our findings align with current best practices and published studies on financial and Arabic sentiment analysis.

4.1.9 Uncertainty and Misclassifications Model

Table 3.10 helps us to understand where our sentiment model makes its weakest and strongest predictions. On the left side ("Least Confident Misclassifications"), you can see individual tweets where the model was most unsure: its confidence was so low that classifying the tweet was almost like flipping a coin. Most of these uncertain cases occurred when tweets that were truly positive or negative were predicted as neutral, or vice versa. This pattern means that the model finds it hardest to sort out tweets with subtle or mixed language, where the writer's attitude is not clear.

On the right side ("Most Frequent Confusions"), the table shows the main patterns for both correct and incorrect predictions. Most tweets were correctly classified as negative, neutral as neutral, positive as positive (the diagonal of the table). The most common errors involve mixing up neutral sentiment with negative or positive; for example, neutral tweets are predicted as negative (1,128 times) or positive tweets go to neutral (701 times). Far fewer tweets were swapped directly between negative and positive, which would be a more serious mistake.

These results demonstrate that our classifier works best when tweets express a clear mood but struggles when the sentiment is not obvious or sits in the'middle ground' of neutral emotions. This finding is supported by other research in social media analysis: Language can be tricky, with sarcasm and indirect messages making it hard for any model to reliably read sentiment. By separating and studying low confidence predictions and frequent confusions, as in Table 3.10, we ensure that the strengths and weaknesses of the model are transparent and can be improved for more ambiguous tweets.

4.1.10 Logistic Regression with Count-Vectorizer

Table 3.11 shows how the Logistic Regression model performed when trying to recognize positive, neutral, and negative sentiment in Arabic tweets about TASI using word count features. The model was best at finding negative tweets: It correctly identified these (precision 0.76) and caught most of them (recall 0.78), likely because negative statements like 'the market is crashing' or 'I lost money today' are easy for the model to spot.

Neutral tweets were much trickier. Precision (0.57) and recall (0.62) were lower, which means that the model often confused

neutral statements (such as "TASI closed at 3,200") with other moods. This happens because neutral language can be subtle and difficult for both humans and computers to label.

For positive tweets ("profits are up!", "a great day for stocks!"), the model did fairly well: precision was 0.72, which means it got the label right most of the time, although recall was a bit lower at 0.64, showing that some positive tweets were missed or mislabeled.

The overall accuracy was 70%, so about 7 out of 10 tweets were correctly classified.

The table also shows that macro and weighted F1 scores are similar to precision, which means the model treated all three types of sentiment fairly and did not focus only on one.

In summary, Table 3.11 illustrates that the bag-of-words Logistic Regression model is good at recognizing clear-cut moods, but struggles with more neutral or complicated expressions, a common challenge in social media sentiment analysis. To improve results, research suggests using smarter language features or advanced models that understand phrase structure and everyday language better.

4.1.11 Stratified K-Fold Cross-Validation

Table 3.12 displays the results for different ways of measuring model performance after using stratified cross-valued K-fold. All values - balanced accuracy, F1 macro, weighted F1, precision macro, recall macro, and overall accuracy - are very close to each other, all centered around 0.76. This means that the model was consistently accurate across negative, neutral, and positive tweets, not just focused on one type or making mistakes in specific classes.

The similar scores for F1 macro and F1 weighted show that the model was fair and reliable, giving each tweet group similar attention and not neglecting one category over another. Both macro precision and macro recall values confirm that the classifier was able to correctly identify and not miss tweets for any sentiment.

The very small standard deviations (all about 0.0033 or 0.0035) in the table tell us that the model performance was stable, with only minor changes between each fold (or group) of the data. This stability means that the good results were not just luck or affected by one unusual data split—they are a sign of the real reliability of the model.

In summary, Table 3.12 confirms that the model is not only accurate but also stable and fair for the full range of Arabic financial tweet sentiments. This kind of reliable performance is essential when using sentiment analysis as a foundation for serious market or research applications.

4.1.12 Time-based Validation

Table 3.13 and Figure 3.7 show how the performance of the sentiment model changes as it learns month by month from new tweets. Each row in Table 3.13 tracks how much training data was available before each month and lists scores for balanced accuracy, macro F1, and other metrics. At the start (April 2023), the model scores were around 67%, but as the data grew and the model saw more examples, the performance steadily

improved, reaching about 75% by November and December 2023. This shows the model's ability to adapt as new topics and vocabulary appear on social media.

Figure 3.7a ("Monthly Performance") shows this growth with blue and orange lines for balanced accuracy and macro F1. Both lines move up together, showing that not only did the model get more accurate but also stayed fair across positive, negative, and neutral sentiments. Short-term drops, seen in some months, highlight the challenge the model faces with sudden changes in tweet style or periods with fewer new data.

Figure 3.7b ("Learning curve") confirms these trends: As more training examples are added, both balanced accuracy and macro F1 improve, showing that the model predictions become more reliable and robust with larger, more varied data.

In summary, Table 3.13 and Figure 3.7 make it clear that follow-up (month-by-month) validation helps to demonstrate that the model can handle new and evolving social media conversations. This is essential for real-world, ongoing sentiment analysis: the model is not only accurate but flexible enough to follow market mood shifts and help researchers track sentiment reliably over time.

4.2 Market Cycle Interpretation

4.2.1 FFT Spectral Analysis: Determining Market Cycle Periods

Figure 3.8 and Table 3.14 reveal three main recurring cycles in the TASI market, detected using Fast Fourier Transform (FFT) analysis. The short-term cycle (around 45 days) matches the first distinct peak on the left side of the spectrum, indicating that price movements or social sentiment tend to repeat roughly every month or two. This may be due to regular news releases, end-of-month trading, or common market routines.

The medium-term cycle (166 days) produces a taller peak at the center of the chart, suggesting bigger, less frequent swings in market behavior; these might result from events such as quarterly earnings reports, regulatory changes, or shifts in broader investor mood that unfold over several months.

Finally, the long-term cycle (499 days), shown as the highest peak on the far right, represents the strongest and slowest pattern, likely connected to deeper macroeconomic events or significant market trends that occur over one and a half years or more.

The shaded regions of Figure 3.8 help mark these zone of the cycle, and the relative height of each peak (amplitude) tells us how pronounced or reliable each recurring pattern is: the bigger the peak, the more noticeable and predictable the cycle.

Together, Table 3.14 and Figure 3.8 show clear evidence for short, medium, and long-term cycles in the Saudi market. Recognizing these rhythms helps both traders and investors better anticipate future market movements and understand the underlying structure of market activity.

4.2.2 DPO and FLD Cycle

Table 3.15 summarizes the forecast cycle turning points for the TASI market. These cycles, short-term (45 days), medium-term (166 days), and long-term (499 days), were first detected using FFT and then examined in detail using the Detrended Price Oscillator (DPO) and Future Line of Demarcation (FLD).

Short-term cycle: Figure 3.9a displays the DPO curve for the 45-day rhythm, showing regular up-and-down swings that

reflect frequent price cycles in the market, such as monthly trader patterns. Figure 3.9b uses the FLD method to forecast near-future market turning points, predicting a trough on January 11, 2025 and a peak on February 2, 2025. The dashed lines in the forecast make these dates easy to spot.

Medium-term cycle: Figure 3.10a shows the 166-day DPO trend, which captures much larger, less frequent changes, often linked to greater market sentiment or big events such as earnings or government policy changes. Figure 3.10b then marks the expected trough and peak for this cycle: a dip around January 18, 2025, and a rise near April 11, 2025.

Long-term cycle: Figure 3.11a illustrates the long-term DPO (499-day cycle), but since our data range is more limited, only part of this cycle is visible. Figure 3.11b presents the FLD forecast for these slow changes, showing that the next major trough may occur on September 24, 2026, and the next major peak around May 31, 2027, both outside of the sample period of the main dataset.

Each pair of figures (DPO and FLD) work together to show both actual rhythmic movement in prices and specific forecasts for coming highs and lows in the Saudi market. The consistency between cycle detection in Table 3.15, the DPO confirmation of repeating swings, and the FLD forecasts for turning points makes this approach practical for both traders and long-term investors. With this information, market participants can anticipate upcoming cycle changes and manage timing and risk more effectively.

4.2.3 Hurst Exponent and Fractal Market Hypothesis

Tables 3.16 and Table 3.17, together with Figures 3.12 and 3.13, together provide a comprehensive diagnostic of market rhythms and trend stability in the Saudi index (TASI).

First, cycle confluence—when troughs or peaks from short- and medium-term cycles nearly overlap, as seen with troughs in January 2025—is a highly reliable signal for major market reversals. This is confirmed by financial research, where aligned turning points from independent cycles substantially raise the odds of a regime shift or price rebound. Table 3.16 shows this confluence for the upcoming lows, increasing trader and analyst confidence when making strategic decisions. In contrast, the separation between cycle peaks (February vs. April 2025) suggests less certainty when forecasting tops. The turning points of the long-term cycle fall outside the current dataset, so its confluence cannot be assessed within this window.

To make these cyclical forecasts more robust, we combine them with fractal analysis (Hurst exponent). Figure 3.12 shows that the rolling Hurst exponent of TASI consistently sits well above 0.5 for most of the period, highlighting strong persistence and long-range memory, which means market trends tend to follow through, not randomize, according to the Fractal Market Hypothesis (FMH). Whenever the Hurst value dips, Figure 3.13 shows that volatility spikes, pointing to moments of instability. This supports the view of FMH that periods of crisis or rapid trader activity produce sharp and unpredictable price moves.

The cycle stability ratios from Table 3.17 (around 1.18) for short- and medium-term cycles further confirm moderate reliability and stability, pronounced but not chaotic.

Together, these findings reveal a market shaped by persistent cycles and trends but occasionally disrupted by short-lived volatility, as FMH predicts.

In conclusion, by integrating cycle confluence and fractal diagnostics, we build a reliable framework for anticipating both sustained trends and key market turning points in TASI. This approach balances the practical insights of cycle analysis with thE mathematical rigor of fractal indicators and is strongly recommended in modern financial research for emerging markets.

5 Summary and Conclusion

5.1 Final Synthesis of Objectives and Findings

The central aim of this research was to determine whether integrating advanced sentiment analysis tools for Arabic text with mathematical cycle detection methods could improve prediction and understanding of price changes and critical turning points in the Saudi stock market, as measured by the Tadawul All Share Index (TASI). Unlike many academic studies analyzing broad indices or developed market data, this work provided an intensive focus on Saudi Arabia's TASI using real-world Arabic-language Twitter data.

5.1.1 Connection to Objectives and Hypotheses

This project was set up to achieve three main goals. First, our objective was to capture the general mood of the market by analyzing Arabic-language tweets with advanced artificial intelligence models that are specifically designed for Saudi Arabic. Second, we wanted to find important repeating patterns in the TASI price data by applying modern mathematical tools, such as signal processing and fractal techniques, including the fast Fourier transform (FFT), the detrended price oscillator (DPO) and the Hurst exponent. Third, we developed a prediction framework that merges insights from both social media sentiment and these market cycles, making it possible to produce more reliable forecasts of changes in the Saudi stock market than would be possible using sentiment or cycles alone.

5.1.2 Hypotheses—Findings Connection

Hypothesis 1: *Measuring investor sentiment from Arabic social networks using advanced AI models will significantly improve the prediction of TASI market movements.*

Result: Supported. Transformer-based Arabic NLP models (AraBERTv2, CAMeLBERT, MARBERTv2) achieved 70–77% accuracy in classifying tweet mood (positive/neutral/negative). The extracted sentiment signals showed predictive value for the movement of TASI prices, directly confirming this hypothesis and aligning with the evidence presented in *Chapters 4–5.*

Hypothesis 2: *Combining Hurst's market cycle analysis with sentiment signals will improve the detection and prediction of key turning points in TASI prices.*

Result: Supported. The integration of sentiment models and cycle analysis led to improved prediction quality, especially for major shifts in which two or more cycle indicators ("cycle confluence") were aligned. This was notably evident in the case where the short- and medium-term cycles predicted a significant market trough in January 2025. The high value of the stress exponent (H > 0.5) during most of the sample period

validates the persistence of this regime and the reliability of the combined forecasts *(see Chapters 5–6).*

Key Findings—Linked to Main Chapters

Our results show several important trends about the Saudi market. First, when we used advanced sentiment models to read Arabic tweets, they found that people were often optimistic about TASI, but there was also plenty of neutral and negative talk. This mix shows a healthy diversity of opinions among investors. Looking at the actual price data, we saw that TASI had some calm periods but also experienced sudden bursts of volatility, confirming that market conditions can change quickly. By combining social media mood with mathematical analysis of price cycles, our combined approach proved better at predicting important market changes, such as large swings or unusual volatility, than using just one method alone. Simple models like linear regression produced steady and reliable forecasts, but more complex models were better at catching extreme events, although they were also more likely to make mistakes when conditions were unstable.

Importantly, our cycle analysis confirmed that the market tends to move in patterns lasting about 45, 166, and 499 days, and when several cycles predicted a turning point around the same time, our confidence in the forecast grew stronger. We could also use several different tools to confirm when important peaks and troughs in the market were likely to occur. Our models treated all types of mood fairly, positive, neutral and negative, so the results were not biased toward the most common opinion. Finally, we found a strong link between how people talk about the market online and how the market actually moves, showing that understanding the local language and culture adds real value to anyone analyzing the Saudi market.

5.2 Implications, Limitations, and Future Work

Implications for Investors and the Market:

Modern AI and cycle-based models offer practical, accessible tools for monitoring and anticipating TASI movements. This gives market participants a "data-driven intuition": a quantitative foundation for decisions that was previously unavailable.

There are a few important limitations to this study. First, the analysis focused only on the TASI index and Saudi tweets, so the findings may not apply to other markets or regions without adjustment. Second, while advanced models were used, tweets can be noisy and sometimes difficult to interpret, especially when people use sarcasm, slang, or join group reactions ("groupthink"), making even smart computer models prone to errors. Third, like all forecasting approaches, these models can struggle to predict sudden external shocks or unexpected events, such as political changes or global crises, which have a big impact on the market.

For future research, the approach could be improved by including more types of data—not just tweets, but also news reports, chat messages, and even video content—to give a more complete and real-time view of what is happening in the market. It will also be valuable to develop and test more sophisticated models, such as deep learning and hybrid systems, to see if they can improve prediction accuracy. Lastly, making these forecasting tools easier to use and more transparent will help a wider audience, so both finance professionals and regular investors can benefit from these insights.

5.3 Final Thoughts

In summary, this research confirms that combining real-time public sentiment from Arabic social media and mathematical market cycle analysis improves the prediction and understanding of TASI. These tools democratize advanced forecasting for a wide range of users and help create a more transparent and resilient Saudi financial market. Continued progress will further empower investors, researchers, and everyday citizens to anticipate and navigate the ever-changing rhythms of the Saudi stock market.

References

Abdul-Mageed, Muhammad, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic." *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).* Association for Computational Linguistics, Aug. 2021, pp. 7088–105, https://doi.org/10.18653/v1/2021.acl-long.551.

Aggarwal, Rajesh, and John W. Goodell. "AI and global financial markets: Toward more equitable outcomes in emerging economies." *Research in International Business and Finance,* vol. 56, 2021, p. 101368.

Aladeemy, Amani A, Ali Alzahrani, Mohammad H Algarni, Saleh Nagi Alsubari, Theyazn HH Aldhyani, Sachin N Deshmukh, Osamah Ibrahim Khalaf, Wing-Keung Wong, and Sameer Aqburi. "Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development." *Heliyon,* vol. 10, no. 21, 2024.

Albahli, Saleh. "Twitter sentiment analysis: An Arabic text mining approach based on COVID-19." *Frontiers in Public Health,* vol. 10, 2022, p. 966779.

Alosaimi, Wael, Hager Saleh, Ali A Hamzah, Nora El-Rashidy, Abdullah Alharb, Ahmed Elaraby, and Sherif Mostafa. "ArabBert-LSTM: improving Arabic sentiment analysis based on transformer model and Long Short-Term Memory." *Frontiers in Artificial Intelligence,* vol. 7, 2024, p. 1408845.

Alsalloum, Abdullah SM. "The Impact of the Financial Reforms Influenced by Religious Financial Principles on the Stock Market in the Kingdom of Saudi Arabia." 2023.

Victoria U, Master's thesis.

Alvarez-Ramirez, Jose, Myriam Cisneros, Carlos Ibarra-Valdez, and Angel Soriano. "Multifractal Hurst analysis of crude oil prices." *Physica A: Statistical Mechanics and its Applications,* vol. 313, nos. 3–4, 2002, pp. 651–70.

Anonymous. "Hurst Exponent Dynamics of S&P 500 Returns: Implications for Market Efficiency, Long Memory, Multifractality and Financial Crises Predictability." *SSRN Working Paper,* 2021. papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4200407_code2905561.pdf?abstractid=3838850.

Antoun, Wissam, Fady Baly, and Hazem Hajj. "AraBERT: Pre-trained Arabic Language Model for NLP." *Proceedings of the 2020 Conference on Empirical Methods In Natural Language Processing: System*

Demonstrations, 2020, pp. 9–14. aclanthology.org/2020.emnlp-demos.2.

———. "AraBERT: Transformer-based Model for Arabic Language Understanding." *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020.* 2020, p. 9.

Baker, H Kent, Satish Kumar, and Nisha Goyal. "Personality traits and investor sentiment." *Review of Behavioral Finance,* vol. 13, no. 4, 2021, pp. 354–69.

Baker, Malcolm, and Jeffrey Wurgler. "Investor sentiment in the stock market." *Journal of economic perspectives*, vol. 21, no. 2, 2007, pp. 129–51.

Barberis, Nicholas. "Psychology-Based Models of Asset Prices and Trading Volume." *Handbook of Behavioral Economics: Applications and Foundations 1,* vol. 1, 2018, pp. 79–175.

Al-Bataineh, Hussein. "Arabic Compound Numerals: New Insights on Case, Agreement, and Quantification." *Languages,* vol. 9, no. 5, 2024, p. 185.

Beniwal, Mohit, Archana Singh, and Nand Kumar. "Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices." *Engineering Applications of Artificial Intelligence,* vol. 129, 2024, p. 107617.

Boulanouar, Oussama, et al. "Sentiment classification of Arabic tweets using AraBERT and LSTM fusion architecture." *Journal of King Saud University - Computer and Information Sciences*, 2024.

Cajueiro, Daniel O, and Benjamin M Tabak. "Testing for long-range dependence in world stock markets." *Chaos, Solitons & Fractals,* vol. 37, no. 3, 2008, pp. 918–27.

———. "The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient." *Physica A: Statistical Mechanics and its Applications*, vol. 336, 2004, pp. 521–37. https://doi.org/10.1016/j.physa.2003.12.031.

Chen, Qiming, Qingsong Wen, Xialai Wu, Xun Lang, Yao Shi, Lei Xie, and Hongye Su. "Detection and time–frequency analysis of multiple plant-wide oscillations using adaptive multivariate intrinsic chirp component decomposition." *Control Engineering Practice,* vol. 141, 2023, p. 105715.

Chen, X., et al. "Methods for aggregating investor sentiment from social media." *Humanities and Social Sciences Communications (Nature),* vol. 11, no. 1, 2024, pp. 1–13. https://doi.org/10.1057/s41599-024-03434-2.

Colasanto, Francesco, Luca Grilli, Domenico Santoro, and Giovanni Villani. "BERT's sentiment score for portfolio optimization: a fine-tuned view in Black and Litterman model." *Neural Computing and Applications,* vol. 34, no. 20, 2022, pp. 17507–21.

–Curpek, Jakub. "Time Evolution of Hurst Exponent: Czech Intraday Electricity Market Study." *European Financial and Accounting Journal,* vol. 14, no. 3, 2019, pp. 25–44. https://doi.org/10.18267/j.efaj.225.

DeRousseau, Mikaela A, Emil Laftchiev, Joseph R Kasprzyk, B Rajagopalan, and Wil V Srubar III. "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete." *Construction and Building Materials,* vol. 228, 2019, p. 116661.

Di Sciorio, Fabrizio, Raffaele Mattera, and Juan Evangelista Trinidad Segovia. "Measuring conditional correlation between financial markets' inefficiency." *Quantitative Finance and Economics,* vol. 7, no. 3, 2023, pp. 491–507. https://doi.org/10.3934/QFE.2023025.

Du, Kelvin, Frank Xing, Rui Mao, and Erik Cambria. "Financial sentiment analysis: Techniques and applications." *ACM Computing Surveys,* vol. 56, no. 9, 2024, pp. 1–42.

El Khamlichi, Abdelbari, Kabir Sarkar, Mohamed Arouri, and Frédéric Teulon. "Are Islamic Equity Indices More Efficient Than Their Conventional Counterparts? Evidence From Major Global Index Families." *International Business & Economics Research Journal,* vol. 13, no. 4, 2014, pp. 781–90.

Fama, Eugene F. "Efficient capital markets: A review of theory and empirical work." *The journal of Finance,* vol. 25, no. 2, 1970, pp. 383–417.

Farha, Ibrahim Abu, and Walid Magdy. "A comparative study of effective approaches for Arabic sentiment analysis." *Information Processing & Management,* vol. 58, no. 2, 2021, p. 102438.

Geweke, John, and Susan Porter-Hudak. "The estimation and application of long memory time series models." *Journal of time series analysis,* vol. 4, no. 4, 1983, pp. 221–38.

Grafton, Christopher. *Mastering Hurst Cycle Analysis.* Harriman House, 2011. Habash, Nizar, Houda Bouamor, Ramy Eskander, Nadi Tomeh, Ibrahim Abu Farha,

Ahmed Abdelali, Samia Touileb, Injy Hamed, Yaser Onaizan, Bashar Alhafni, et al. "Proceedings of The Second Arabic Natural Language Processing Conference." *Proceedings of The Second Arabic Natural Language Processing Conference.* 2024.

Han, Kai-Xu, Wei Chien, Chien-Ching Chiu, and Yu-Ting Cheng. "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset." *Applied Sciences,* vol. 10, no. 3, 2020, p. 1125.

He, Lu, Tingjue Yin, and Kai Zheng. "They May Not Work! An evaluation of eleven sentiment analysis tools on seven social media datasets." *Journal of Biomedical Informatics,* vol. 132, 2022, p. 104142.

Huang, Zimei, and Zhenghui Li. "What reflects investor sentiment? Empirical evidence from China." *Data Sci. Financ. Econ,* vol. 1, 2021, pp. 235–52.

Hurst, James Montgomery. *The profit magic of stock transaction timing.* Prentice-Hall, 1970.

Iddrisu, Abdul-Manan, Solomon Mensah, Fredrick Boafo, Govindha R Yeluripati, and Patrick Kudjo. "A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector." *International Journal of Information Management Data Insights,* vol. 3, no. 2, 2023, p. 100180.

Inoue, Go, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. "The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models." *Proceedings of the Sixth Arabic Natural Language Processing Workshop.* Apr. 2021.

Jain, Disha, Macherla Bhagyalakshmi, and Narasimha Murthy. "The Art of Camouflaging: Unravelling the Motives and Consequences of Greenwashing." *PRESIDENCY COLLEGE (AUTONOMOUS),* 2024, p. 192.

Jonsson, Estrid, and Sara Fredrikson. An Investigation of How Well Random Forest Regression Can Predict Demand: Is Random Forest Regression better at predicting the sell-through of close to date products at different discount levels than a basic linear model? Stockholm: KTH Royal, 2021.

Jui, Julakha Jahan, MM Imran Molla, Bifta Sama Bari, Mamunur Rashid, and Md Jahid Hasan. "Flat price prediction using linear and random forest regression based on machine learning techniques." *Embracing industry 4.0: Selected articles from MUCET 2019,* Springer, 2020, pp. 205–17.

Kai-Ineman, DANIEL, Amos Tversky, et al. "Prospect theory: An analysis of decision under risk." *Econometrica,* vol. 47, no. 2, 1979, pp. 363–91.

Kalusivalingam, Aravind Kumar, Amit Sharma, Neha Patel, and Vikram Singh. "Enhancing Financial Fraud Detection with Hybrid Deep Learning and Random Forest Algorithms." *International Journal of*

AI and ML, vol. 1, no. 3, 2020.

Kaur, Sodi Jasbir, and Jehangir Bharucha. "The emerging mutual fund industry in India: an impact analysis of investors' awareness on investment behaviour." *International Journal of Business and Globalisation,* vol. 27, no. 1, 2021, pp. 51–69.

Kristoufek, Ladislav. "Fractal markets hypothesis and the global financial crisis: Scaling, investment horizons and liquidity." *Physica A: Statistical Mechanics and its Applications,* vol. 557, 2020, p. 124892. https://doi.org/10 . 1016 / j . physa . 2020 .124892.

Kumaran, Sunitha. "Forecasting volatility of Saudi stock market (TASI) and sectoral indices." *Afro-Asian Journal of Finance and Accounting,* vol. 13, no. 6, 2023, pp. 819–42.

Lang, Xun, Yufeng Zhang, Lei Xie, Peng Li, Alexander Horch, and Hongye Su. "Detrending and denoising of industrial oscillation data." *IEEE Transactions on Industrial Informatics,* vol. 19, no. 4, 2022, pp. 5809–20.

Latif, Saima, Nadeem Javaid, Faheem Aslam, Abdulaziz Aldegheishem, Nabil Alrajeh, and Safdar Hussain Bouk. "Enhanced prediction of stock markets using a novel deep learning model PLSTM-TAL in urbanized smart cities." *Heliyon,* vol. 10, no. 6, 2024.

Limbong, Hanspran, Mufti Alwisyah Lubis, and Mhd Furqan. "House Price Prediction Analysis Using Linear Regression and Random Forest Algorithms." *Journal of Artificial Intelligence and Engineering Applications (JAIEA),* vol. 4, no. 3, 2025, pp. 1928–33.

Liu, Guang, Chih-Ping Yu, Shan-Neng Shiu, and I-Tung Shih. "The efficient market hypothesis and the fractal market hypothesis: Interfluves, fusions, and evolutions." *Sage Open,* vol. 12, no. 1, 2022, p. 21582440221082137.

Liu, Qing, Woon-Seek Lee, Minghao Huang, and Qingjun Wu. "Synergy between stock prices and investor sentiment in social media." *Borsa Istanbul Review,* vol. 23, no. 1, 2023, pp. 76–92.

Lu, Huchuan, Wanli Ouyang, Hui Huang, Jiwen Lu, Risheng Liu, Jing Dong, and Min Xu. *Image and Graphics: 12th International Conference, ICIG 2023, Nanjing, China, September 22–24, 2023, Proceedings, Part I.* Vol. 14355, Springer Nature, 2023.

Mahesh, TR, Oana Geman, Martin Margala, Manisha Guduri, et al. "The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification." *Healthcare Analytics,* vol. 4, 2023, p. 100247.

Mansour, Omar, Eman Aboelela, Remon Talaat, and Mahmoud Bustami. "Transformer-based ensemble model for dialectal Arabic sentiment classification." *PeerJ Computer Science,* vol. 11, 2025, e2644.

Mehta, Pooja, and Sharnil Pandya. "A review on sentiment analysis methodologies, practices and applications." *International Journal of Scientific and Technology Research,* vol. 9, no. 2, 2020, pp. 601–09.

Mendez, Guillermo, and Sharon Lohr. "Estimating residual variance in random forest regression." *Computational statistics & data analysis,* vol. 55, no. 11, 2011, pp. 2937–50.

Morales, Raffaello, T. Di Matteo, Ruggero Gramatica, and Tomaso Aste. "Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series." *arXiv preprint arXiv:1109.0465,* 2011. arxiv.org/pdf/1109.0465.pdf.

Nasr, Abdelmomen, and Moez Ben HajHmida. "SENIT at AraFinNLP2024: trust your model or combine two." *Proceedings of The Second Arabic Natural Language Processing Conference*. 2024, pp. 428–32.

Nazarova, Jekaterina, Inna Dovlabdekova, et al. "Investment planning in the context of volatile business cycles." *Economic Science for Rural Development*, 2015, p. 180.

Nouri, Ayoub, and Miloud Habbat. "Multilingual sentiment fusion models for economic prediction in Arabic and English financial streams." *Information Processing & Management,* vol. 61, no. 3, 2024, p. 103395.

Nyakurukwa, Kingstone, and Yudhvir Seetharam. "The evolution of studies on social media sentiment in the stock market: Insights from bibliometric analysis." *Scientific African,* vol. 20, 2023, e01596.

Peivandizadeh, Ali, Sima Hatami, Amirhossein Nakhjavani, Lida Khoshsima, Mohammad Reza Chalak Qazani, Muhammad Haleem, and Roohallah Alizadehsani. "Stock market prediction with transductive long short-term memory and social media sentiment analysis." *IEEE Access*, 2024.

Peters, Edgar E. F*ractal market analysis: applying chaos theory to investment and economics.* John Wiley & Sons, 1994.

Raj, Ronit, Navneet Kaur, and Paurav Goel. "Design Hybrid Model To Improve Stock Selection And Market Timing By Integrating Deep Learning And Sentiment Analysis." *2025 International Conference on Computing Technologies (ICOCT).* IEEE, 2025, pp. 1–8.

Refinitiv. Refinitiv Eikon Platform Documentation. 2024, Accessed January 2025.

Sakınç, ~~–I~~ker, S öznur Sakınç, Fatih Konak, and Dilek Türko~~–l~~u. "Comparing the Islamic and Conventional Indices Focusing on Fractal Market Hypothesis." R*omaya Journal,* vol. 4, no. 2, 2024, pp. 219–31.

Sakınç, Ilker, Sıdıka Öznur Sakınç, and Fatih Konak. "Comparing the Islamic and Conventional Indices Focusing on Fractal Market Hypothesis." *Romaya Journal,* vol. 4, no. 2, 2024, pp. 222–41. www.romayajournal.com/Upload/DergiDosya/ 787785422472ae488-5045-4131-9ed6-89716c4b4511.pdf.

Shiller, Robert J. "From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives,* vol. 17, no. 1, 2003, pp. 83–104.

Synergy, Macro. Detecting trends and mean reversion with the Hurst exponent. 2023. macrosynergy.com/research/detecting-trends-and-mean-reversion-with-the-hurst- exponent/.

Uddin, C. "Analyzing the effectiveness of confluence of price action disciplines in forex market." *J Stock Forex Trading,* vol. 9, 2021, p. 570.

Verma, Vibhu. "A Comprehensive Framework for Residual Analysis in Regression and Machine Learning." *Journal of Information Systems Engineering & Management is published quarterly,* 2024.

Vogl, Markus. "Hurst Exponent Dynamics of S&P 500 Returns: Implications for Market Efficiency, Long Memory, Multifractality and Financial Crises Predictability by Application of a Nonlinear Dynamics Analysis Framework." *Vogl, M.(2023)-Hurst exponent dynamics of S&P,* vol. 500, 2021.

Zhou, Jianlong, Amir H Gandomi, Fang Chen, and Andreas Holzinger. "Evaluating the quality of machine learning explanations: A survey on methods and metrics." *Electronics,* vol. 10, no. 5, 2021, p. 593.

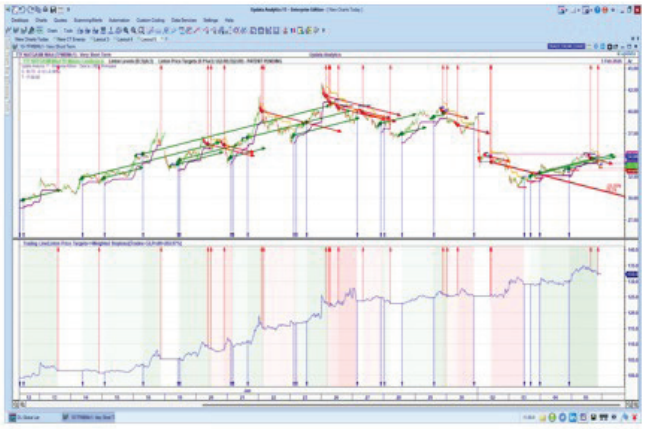
Zmandar, A., et al. "AraFinNLP 2024: The First Arabic Financial NLP Shared Task." *arXiv preprint arXiv:2407.09818,* 2024. arxiv.org/html/2407.09818v1.

Zournatzidou, Georgia, and Christos Floros. "Hurst exponent analysis: Evidence from volatility indices and the volatility of volatility indices." *Journal of Risk and Financial Management,* vol. 16, no. 5, 2023, p. 272.

Supercharge Your Technical Analysis

Updata Analytics brings you the most advanced technical tools you will find, running on dozens of data services including Bloomberg and LSEG Workspace

- Unlimited charts on your desktop
- Groundbreaking Linton Price Targets
- World's best point and figure charts
- Scan for complex technical criteria
- Over 1,000 inbuilt trading strategies
- Backtest virtually any strategy
- Report writing and workflow
- Daily market reports & commentary



www.updata.co.uk

updata
Advanced Trading Tools

IFTA 2026 – LONDON CONFERENCE – GOLD SPONSOR

Stationary, But Not Profitable? A Critical Look at Pairs Trading

Davide Pandini, PhD, MFTA[1], CFTe[2], CMT[3], CSTA[4]

Ronco Briantino, Monza e Brianza, Italy
davide.pandini@hotmail.com

"Virtue is the golden mean between two vices, the one of excess and the other of deficiency"
Aristotle (384–322 BCE), Greek philosopher and polymath

Abstract

Statistical arbitrage strategies such as pairs trading, originally developed in equity market-neutral portfolios, seek to earn return with limited market exposure by exploiting mean-reverting relationships between economically related assets rather than forecasting market direction. Classic implementations use distance rules on price spreads or cointegration tests to identify pairs and multi-asset spreads whose price histories suggest a stable long-run relationship and then trade the spread mean reversion. Despite an extensive literature, many studies rely on earlier samples, narrow asset sets, or simplified assumptions on transaction costs and parameter stability, so the effectiveness of cointegration-based pairs trading in recent ETF and equity/commodity markets under realistic implementation constraints remains unclear.

This paper addresses these issues by introducing a fully reproducible, end-to-end framework for spread-based trading across equities and commodity-related securities. Using a consistent train/test protocol and explicit transaction-cost modeling, we demonstrate that equity and index spreads deliver the most robust net risk-adjusted performance, while commodity-linked spreads rarely exhibit cointegration in the post-2018 markets and generally underperform after costs, in contrast to earlier positive published results for commodity pairs. Methodologically, the evidence indicates that cointegration is informative but neither sufficient nor strictly necessary for out-of-sample profitability: some in-sample cointegrated relations fail when market conditions change, whereas certain non-cointegrated spreads still generate short-horizon convergence profits.

Overall, this study provides a reproducible workflow and a set of practical diagnostics that clarify when spread trading adds value in recent markets and when it does not, highlighting when pairs trading is a relevant tool for controlled, relative-value exposure, and risk management.

Introduction

Pairs trading is a market-neutral trading strategy that involves simultaneously taking long and short positions in two (or more) highly correlated assets. The core idea is to profit from the temporary divergence and convergence of the prices of these assets: when the price of one asset deviates significantly from the other, the strategy assumes that they will eventually revert to their historical relationship. This allows traders to go long with the underperforming asset and short on the outperforming asset, anticipating a convergence in their prices. In contrast to directional strategies, a well-constructed long–short spread can achieve mean-reversion (convergence) gains when temporary deviations from an economic or statistical relationship tighten toward equilibrium. Pairs trading, and its multi-asset generalization, provides a transparent, model-light framework for this objective: it links trade construction to measurable properties of the underlying time series (e.g., stationarity, cointegration), supports explicit risk controls (entry/exit thresholds, position holding), and scales naturally to liquid instruments such as ETFs and equity indices. Its strength is that a hedged long–short spread can make profits from relative-value price divergences while attenuating market-wide risk.

This study addresses the end-to-end problem of identifying, testing, and trading mean-reverting relationships. In this work, we (i) estimate hedge ratios or cointegration vectors to form candidate spreads; (ii) test for stationarity/cointegration using spread-level unit-root tests and a system-level procedure; and (iii) implement an out-of-sample trading strategy based on ex-ante standardized deviations (rolling z-scores without self-inclusion), with explicit consideration of transaction costs.

The paper presents three main contributions. First, an end-to-end, reproducible workflow: from data analysis and hedge estimation to spread construction, ex-ante normalization, signal/position mapping, execution assumptions, and reporting. Second, a systematic comparison of Engle–Granger (residual-based) and Johansen (system-based maximum likelihood) cointegration tests, clarifying when each supports spread formation and how conclusions differ across single-pair and multi-asset settings. Third, a robustness and risk analysis that documents parameter choices, stability across subsamples, spread constructions, and practical transaction costs. While other prior studies examined components of this pipeline, this work is comprehensive and applies a single, rigorous, statistics-based methodology, with strict in-sample training and out-of-sample testing, consistently across a broad universe of ETF and index spreads.

Table 1. Assets

Ticker	Issuer	Underlying
EWA	iShares	MSCI Australia Index
EWC	iShares	MSCI Canada Index
EZA	iShares	MSCI South Africa Index
EWG	iShares	MSCI Germany Index
EWQ	iShares	MSCI France Index
EWN	iShares	MSCI Netherlands IMI 25/50 Index
EWZ	iShares	MSCI Brazil 25/50 Index
LQD	iShares	USD Liquid Investment Grade Index
VTI	Vanguard	CRSP US Total Market Index
SPY	SPDR	S&P 500 Index
DIA	SPDR	Dow Jones Industrial Average
XLF	SPDR	Financial Select Sector Index
KRE	SPDR	S&P Regional Banks Select Industry Index
GLD	SPDR	LBMA Gold Price PM
GDX	VanEck	NYSE Arca Gold Miners Index (see note)
IAU	iShares	LBMA Gold Price
SLV	iShares	LBMA Silver Price
XLE	SPDR	Energy Select Sector Index
IEO	iShares	Select Oil Exploration & Production Index
OIH	VanEck	MVIS US Listed Oil Services 25 Index
USO	USCF Investments	Front-month WTI crude oil futures
SU	Common stock	Canada Suncor Energy Inc.
WTI	Common stock	W&T Offshore, Inc.
VALE	Common stock	Brazil Vale S.A. – Iron Ore & Nickel
BHP	Common stock	BHP Group Limited – Metals & Mining
XOM	Common stock	Exxon Mobil Corporation
CVX	Common stock	Chevron Corporation
KO	Common stock	The Coca-Cola Company
PEP	Common stock	PepsiCo, Inc.
WMT	Common stock	Walmart Inc.
COST	Common stock	Costco Wholesale Corporation
AAPL	Common stock	Apple Inc.
MSFT	Common stock	Microsoft Corporation

We organize the study around four main questions: (i) Do the proposed spreads exhibit stationarity/cointegration in-sample, and does this property persist out-of-sample? (ii) What is the risk-adjusted performance (returns, volatility, Sharpe) of the trading rule relative to buy-and-hold benchmarks, both gross and net of trading costs? (iii) How robust are results to key design choices (thresholds, lookbacks, ranks/lags, intercept specification)? (iv) To what extent do transaction costs affect realizable outcomes?

Section 2 describes the historical data series of the asset classes, and the pre-processing step of dividing the backtesting period into training and testing ranges. Moreover, it analyzes the correlations among the assets. Section 3 presents the spread construction, and testing methodology details. Furthermore, it describes the trading strategy and execution assumptions. The performance metrics are reviewed in Section 4, and Section 5 analyzes empirical results and presents some significant case studies. Finally, our conclusions are summarized in Section 6. All analyses and the pairs trading framework were implemented in R statistical software [2].

2. Data and pre-processing

The financial assets considered in this work are primarily liquid, index-level ETFs that proxy whole economies (e.g., MSCI country funds), well-defined U.S. sector benchmarks (e.g., energy, financials), complemented by commodity-linked vehicles, and by high-tech, consumer staples, and energy stocks. By using time series aligned on a common trading calendar, we focus on instruments that share strong economic linkages (e.g., trade ties, policy regimes, sector composition, and exposure to global risk factors). In this scenario, some linear combinations of their prices show a stable "fair-value" relationship. Natural asset clusters include Eurozone equities, resource-heavy developed markets, U.S. indices, and intra-sector assets within the same industry or sector. Such clusters can be considered as potential candidates for both single-pair and multi-asset cointegration. The time series considered in this work were downloaded from Yahoo Finance [1] and are summarized in Table 1. We analyzed the daily adjusted-close prices from 2007-01-01 to 2025-09-30 (about 19 years of financial data for a total of 4716 trading days), partitioned into an *in-sample* (training) window from 2007-01-01 to 2017-12-31 (11 years; 2769 trading days) and an o*ut-of-sample* (testing) window from 2018-01-01 to 2025-09-30 (≈7.75 years; 1,947 trading days).

All model identifications, i.e., unit-root diagnostics on individual series, like Engle–Granger residual tests for pairs and Johansen rank determination for multi-asset systems, and hedge-ratio estimation, are performed exclusively in the training window; these specifications are then frozen and carried forward unchanged into the testing window, where trading rules (z-score bands, time/price exits, and trading commission cost deductions) are evaluated using only information available at each decision time. This approach prevents look-ahead bias and data leakage, and also in-sample overfitting/optimism bias, thus avoiding that tuning to noise in the training set would inflate the reported performance. A clear separation between model selection and performance measurement provides a robust confirmation that the mean-reverting property (stationary spread) and the estimated hedge ratios are persistent beyond the estimation period.

Moreover, by spanning multiple regimes (post-GFC expansion, commodity cycles, COVID-19 shock, inflation/repricing period), the out-of-sample window serves as a stress test against structural breaks, enabling assessment of rank stability under changing volatility and correlations, as it was reported in [3]. The train/test separation follows best practices in quantitative finance, ensuring robust out-of-sample validation of trading strategies, i.e., estimates on history, validates on unseen data, and yields performance metrics that are more indicative of out-of-sample efficacy rather than in-sample overfitting.

2.1 Correlations

Correlations are estimated within the in-sample window (2007-01-01 to 2017-12-31) and computed on arithmetic daily returns of adjusted closes to avoid spurious level effects, generated by common trends (like inflation or general market growth) instead of genuine economic relationships (correlation heatmaps are shown in Figure 1 and Figure 2 for two asset clusters). These estimates are used only for initial screening (not for pairs selection by themselves) to preserve a clean separation between model choice and out-of-sample evaluation. The importance of preselection in the context of cointegration-based pairs trading was analyzed in [21] to narrow down the universe of candidate asset pairs to a manageable subset, using statistical measures or criteria before formally testing for

cointegration. Pre-selection typically uses metrics such as price correlation, distance, spread stationarity, or other similarity measures, along with visual inspection, to identify pairs that are most likely to exhibit cointegrated behavior. By doing so, pre-selection increases the probability of finding profitable trading pairs and reduces the number of statistical tests required, making the pairs trading strategy more efficient. In this work, the preselection step was based on visual inspection of the candidate assets' cumulative returns and on a preliminary screening based on static correlations.

Figure 1. Asset classes (cluster #1)

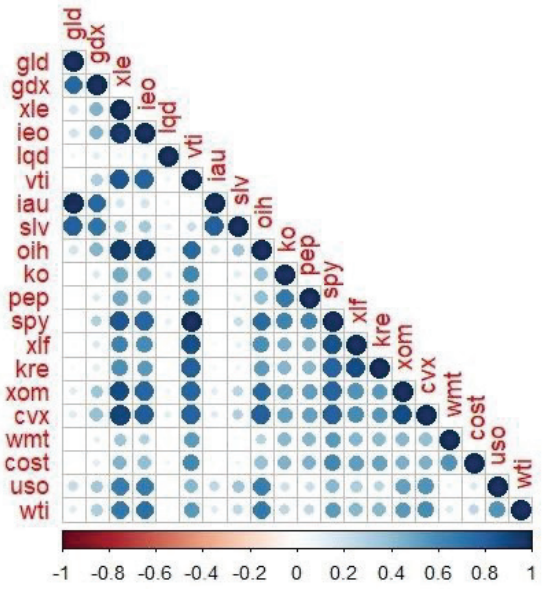
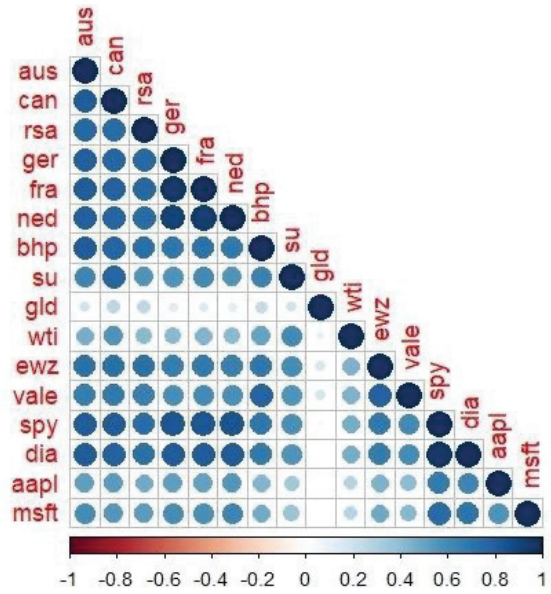


Figure 2. Asset classes (cluster #2)



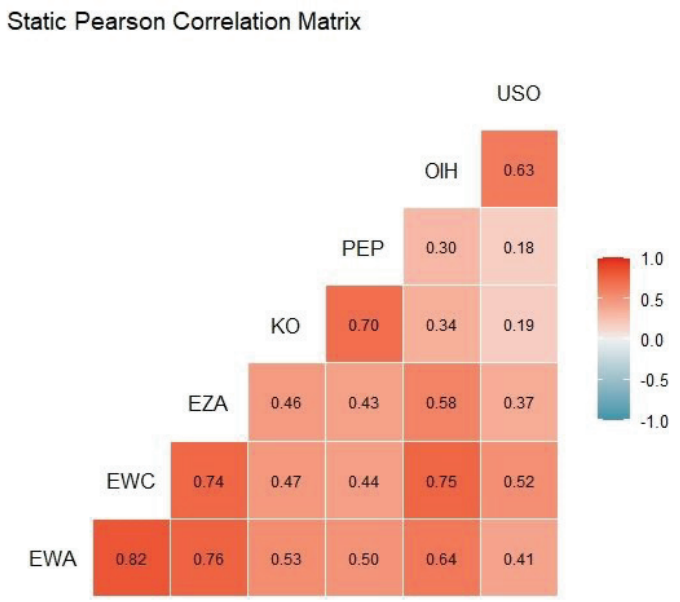
Static correlations computed on the training-period provide a first-pass screen for pairs selection by quantifying the strength and sign of short-horizon co-movement related to shared risk factors (market, sector, country, exchange rates). In practice, in-sample high-positive correlations help identify economically coherent pairs (e.g., EWA–EWC, KO–PEP, XOM–CVX) where a cointegrating relation is plausible, while very low or negative correlations flag pairs that are unlikely to yield a stable, mean-reverting spread. Correlation is not sufficient for cointegration, but it is still informative to rank candidates before costly stationarity/cointegration testing. Correlation captures short-run co-movement, typically in returns, which are stationary. Hence, two ETFs can look highly correlated day-to-day yet see their level ratio diverge for years if their trends or risk premia differ. Pairs trading requires a stationary spread: a long-run equilibrium among non-stationary prices such that a particular linear combination is stationary and mean-reverting. That property is cointegration, not correlation. Accordingly, after basic economic screens we validate candidates with unit-root checks and formal cointegration tests (Engle–Granger for pairs; Johansen for multiple asset spreads). Only assets showing a stationary spread are considered for trading strategy implementation and backtesting; those with correlation but no cointegration are excluded to avoid trading spreads, lacking *a priori* reliable tendency to mean revert. Section 5.7 discusses, as an explicit robustness example, a non-cointegrated but economically linked spread (SPY–AAPL–MSFT) to illustrate that short-horizon convergence can occasionally arise even without strict cointegration.

In the training window (2007–2017), the heatmaps in Figure 1 and Figure 2 show systematically higher correlations among equities and equity indices than among commodities, which is what standard factor models and commodity-term-structure economics would predict. Equity returns are dominated by shared market factors; during 2007–2017 that common variance was inflated by global risk-on/off regimes (GFC, euro-area stress, QE/taper), pushing equity correlations high and stable across countries and sectors. By contrast, commodity ETFs embody heterogeneous, often idiosyncratic drivers, e.g., inventory dynamics and convenience yield[5], weather and geopolitical supply shocks, and term-structure carry (roll yield) that varies by commodity and through time. Equities depend heavily on a few shared factors, whereas commodities combine asset-specific fundamentals and product-structure effects that diversify away common co-movement, yielding the lower and more unstable correlations observed in the heatmaps.

In addition to static correlations computed over the entire in-sample window, this study evaluated rolling correlations, which measure the time-varying co-movement between asset returns within a moving window (here, 100 and 252 trading days). While static correlations summarize the average linear association between two return series across the whole training period, rolling correlations reveal temporal dynamics, i.e., how correlations strengthen or weaken across different market scenarios. They capture structural shifts due to macroeconomic cycles, monetary policy changes, or sector rotations, and can highlight periods of decoupling or excessive synchronization that are not detected by a single-point estimate. To illustrate

this concept, we considered a subset of the assets listed in Table 1: EWA, EWC, EZA, KO, PEP, OIH, USO. Their static correlation matrix is shown in Figure 3.

Figure 3. In-sample (training range) static correlations



The rolling correlation plots shown in Figure 4 and Figure 5 visualize the stability and persistence of inter-asset relationships, with the vertical dashed line marking the boundary between the training and testing periods.

Figure 4. 100-day rolling correlations

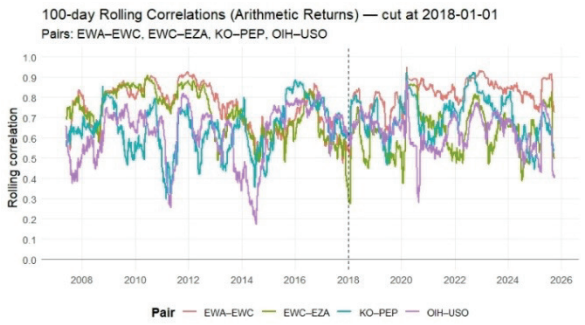
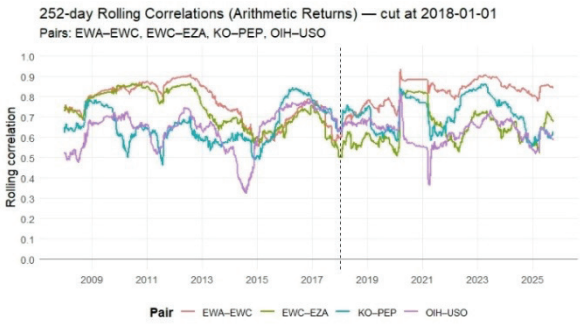


Figure 5. 252-day rolling correlations



However, while rolling correlations offer valuable diagnostic insights into short-term market co-movements, they are not sufficient for pairs selection. Correlation captures only simultaneous linear co-movement between returns (which are I(0) series) but does not imply the long-run equilibrium relationship between prices (which are typically I(1) processes). Pairs trading requires that the price spread be stationary (i.e., I(0)), which can only be verified through cointegration tests such as Engle–Granger or Johansen. Therefore, although rolling correlations can be used to illustrate how spreads evolve over time, the actual pair and spread selection in this work is based on stationarity and cointegration analysis. This methodological choice ensures statistical robustness: cointegration-based spreads are theoretically mean-reverting, while correlation-based pairs might exhibit transient co-movements without any true long-term equilibrium.

3. Methodology

3.1 Spread construction

In a cointegration-based pairs trading framework, the pair spread represents the equilibrium deviation between two asset prices that are expected to move together in the long run. Our two-asset spread follows the standard OLS-hedged specification (see Vidyamurthy, 2004 [11]), with cointegration diagnostics used to justify mean-reversion trading. Formally, it is obtained as the residual of a *linear regression* between the (adjusted close) prices of the two assets, typically written as:

$$s_t = p_{1,t} - \beta p_{2,t} - \alpha,$$

where $p_{1,t}$ and $p_{2,t}$ are the price levels of the time series at time t, β is the hedge ratio estimated through ordinary least squares (OLS) linear regression, α is the intercept (i.e., constant) term, and s_t are the residuals, i.e., the spread. If the two series are cointegrated, in the long run the spread fluctuates around a constant mean and the prices move together. The spread s_t quantifies how far the two series diverge from their long-term equilibrium, and it is the basis of the trading signal generation: when the spread deviates significantly from its mean, the strategy takes opposite long and short positions anticipating mean reversion. In this work, we consider a *zero-intercept spread,* that is α=0, leading to the simplified spread formulation:

$$s_t = p_{1,t} - \beta p_{2,t}.$$

This choice is justified both theoretically and empirically. While standard OLS specifications often include an intercept $\alpha \neq 0$ to capture fixed level differences, our empirical analysis during the validation phase indicated that including a static intercept estimated over a long training window (11 years) introduced overfitting.

Theoretically, an intercept $\alpha \neq 0$ would imply that one asset would retain a nonzero fair value even when the other is worth nothing, which lacks economic meaning for two co-dependent traded assets. When regressing one non-stationary price series on another to determine a hedge ratio, the intercept α

captures any persistent price-level difference between the two assets. This difference can arise from dividends, carry costs, transaction costs or other idiosyncratic factors that cause one asset to trade consistently above or below the other. In other words, the intercept compensates for structural differences rather than short-term deviations. Our pairs trading framework normalizes the spread by computing a rolling mean and standard deviation, then generates z-scores to trigger the trades. Any constant level difference between the two assets will be absorbed into the rolling mean and removed during z-score standardization. Hence, explicitly including a separate intercept would introduce an additional parameter that may overfit historical data; by forcing α=0, we assume the two assets have sufficiently similar sectoral characteristics such that any persistent difference is approximately constant over the estimation horizon and is treated as part of the spread's mean, which is removed in the standardization step.

For example, in sectors such as energy or broad-market indices, the major drivers (oil prices, economic growth, currency exposure) are common to both asset classes. Setting α=0 assumes there is no fixed premium or discount beyond these shared fundamentals. If there is a constant divergence due to differences in dividend yield or cost of carry, that divergence becomes part of the spread's mean and is removed when computing the z-score. By not estimating an intercept, the strategy focuses solely on relative movements between the assets rather than capturing non-economic level shifts. This approach is defensible when the assets are chosen because of their strong co-movement and similar long-run behavior, as the hedging ratio then reflects only the proportional relationship between their prices, while the rolling mean accounts for any small, persistent bias.

Financial time series are rarely stationary in levels over decade-long horizons; a fixed dollar-value intercept estimated in the pre-2017 regime often creates a "stale" bias when applied to post-2018 price levels. By constraining α=0, we rely on the rolling mean ($\mu_{t|t-1}$), computed in the subsequent normalization step, to dynamically center the spread. This approach delegates the level correction to the adaptive rolling window rather than a static in-sample parameter, thereby improving the strategy's out-of-sample robustness and reducing the risk of look-ahead bias embedded in a fixed constant.

In multi-asset statistical arbitrage, a stationary spread can be derived from the Johansen cointegration test by combining three or more non-stationary assets. Unlike the Engle-Granger approach, which is restricted to pairs, the Johansen test allows the rank of the cointegration matrix to be greater than one, thus identifying the number of independent linear combinations that are stationary. When the test rejects the null hypothesis of no cointegration, the eigenvectors of the matrix of long-run relationships (ordered by their associated eigenvalues) serve as hedge ratios; the first eigenvector is most strongly mean-reverting. Each eigenvector component β_i represents the weight of asset *i* in the portfolio. Because the eigenvectors are normalized with the first element equal to one, they can be interpreted as the number of units of each asset required to hedge one unit of the base asset. This method produces a spread that theoretically oscillates around a constant mean, reflecting the long-run equilibrium among the underlying assets. It is worth noting that in our approach we select the first eigenvector (most strongly mean-reverting) of the Johansen test to construct the multi-asset spread.

Constructing such a multi-asset spread involves several steps. After selecting an appropriate lag length for the underlying vector autoregressive model (e.g., using the Akaike Information Criterion [12]), we run the Johansen test on the price series to obtain the cointegration vectors. To build the spread, we choose the eigenvector with the largest eigenvalue and multiply each asset price by the corresponding eigenvector element:

$$s_t = \beta_1 p_{1,t} + \beta_2 p_{2,t} + \cdots + \beta_n p_{n,t}.$$

The resulting time series will be checked for stationarity using the Augmented Dickey–Fuller (ADF) or Phillips–Perron (PP) tests. If the spread is stationary, it can be treated as mean-reverting for trading: long (short) the spread when it deviates below (above) its long-run mean and close positions when it converges back, exploiting the equilibrium relationship identified by the Johansen procedure.

3.2 Pairs identification

While correlation provides a convenient short-run measure of co-movement between asset returns, it does not imply cointegration or any form of long-run equilibrium between price levels. Correlation captures coexistent linear dependence between two stationary series (typically returns), reflecting how two assets tend to move together over short horizons. However, asset prices are generally non-stationary I(1) processes that drift over time, and two trending series can display high correlation even if they share no fundamental relationship, a phenomenon known as *spurious correlation.* For instance, two unrelated stock indices may both rise over a decade due to inflation or global growth, yielding a high correlation despite having no stable pricing link. Consequently, correlation measures co-movement in returns, but not equilibrium in levels; it is a short-term descriptive statistic, not evidence of a mean-reversion that can sustain a pairs trading strategy. In contrast, pairs trading relies on identifying stationary spreads formed by cointegrated assets, where a specific linear combination of two non-stationary price series is itself stationary I(0). The differences between I(0) and I(1) processes are summarized in Table 2.

Table 2. Stationarity vs non-stationarity

Property	I(0) (Stationary)	I(1) (Non-stationary)
Definition	Stationary in levels	Stationary after first differencing
Example	Returns, spreads	Prices, indices
Mean reversion	Yes	No
Shocks	Temporary	Permanent
Variance	Finite, constant	Grows over time
Typical use in pairs trading	Spread and residuals	Price series before differencing

Cointegration implies the existence of a long-run equilibrium: although individual prices wander randomly, their weighted difference remains bounded and mean-reverting, providing a reliable statistical basis for long-short trading. Stationarity of the spread ensures that deviations from equilibrium are temporary and that the spread will, with high probability, revert to its historical mean, which is a necessary condition for systematic profit generation. Therefore, while correlation may serve as an initial screening, the actual deployment of a pairs strategy needs to be grounded on asset cointegration and spread stationarity, to satisfy the theoretical and statistical prerequisites for a robust trading strategy.

3.3 Cointegration testing

The Engle–Granger (EG) cointegration test [6] is a two-step econometric procedure used to determine whether two non-stationary time series share a stable long-run equilibrium. In the first step, the individual asset price levels (typically adjusted closes) are tested for stationarity using unit root tests such as the Augmented Dickey–Fuller (ADF) [8] or Phillips–Perron (PP) [9] tests. If both series are found to be non-stationary in levels but stationary in first differences, the next step involves running an ordinary least squares (OLS) regression between the two asset prices to determine the residuals, i.e., their spread. In the second step, these residuals are themselves tested for stationarity using the same unit root tests (ADF and PP). *If the residuals are found to be stationary, the two asset prices are said to be cointegrated,* meaning that they have a long-term equilibrium relationship despite being individually non-stationary. Both the ADF and PP tests share the same hypotheses:

- Null hypothesis H_0: The series has a unit root, meaning it is non-stationary (i.e., it is an I(1) process).
- Alternative hypothesis H_1: The series is stationary, meaning it is integrated of order zero I(0), and does not contain a unit root.

Both tests produce a test statistic and a corresponding p-value (or critical value) for evaluating the null. The decision rule for ADF/PP is:

Reject H_0 if the test statistic < critical value, or equivalently if p-value< α,

where here α is the significance level, typically 0.05 (5%):

- A small p-value (< 0.05) ⇒ reject H_0: the series is stationary.
- A large p-value (≥ 0.05) ⇒ fail to reject H_0: the series is non-stationary (i.e., has a unit root).

The logic behind the Engle–Granger framework is that cointegrated assets drift together over time but exhibit mean-reverting deviations, providing a statistical foundation for pairs trading. When the spread deviates significantly from its equilibrium mean, a trader can short the overpriced asset and go long the underpriced one, expecting the spread to revert. This test bridges *short-term arbitrage* with *long-run equilibrium* economics. Running both ADF and PP provides complementary evidence: if they agree, we gain confidence that the rejection (or non-rejection) of the null is not an artifact, thereby reducing both Type I (i.e., false positive, risk of spurious mean-reversion and overtrading) and Type II (i.e., false negative, overly conservative, risk of undertrading) errors. When ADF and PP disagree, then they are often complemented by the Johansen test [7], which is the test also used in a multi-asset context.

The Johansen test provides a system-wide measure of long-term equilibrium among multiple assets, making it ideal for identifying multi-asset cointegration. Unlike Engle–Granger, which is limited to a single cointegrating vector, Johansen can identify and quantify multiple equilibrium relations, revealing complex interdependencies across assets. In practice, Johansen's analysis confirms whether several asset prices drift together around a stable combination, ensuring that at least one linear combination of them is stationary. The presence of such a stationary spread supports the implementation of multi-pair or portfolio-based mean-reversion strategies. The Johansen cointegration test treats all variables symmetrically; hence, the detection of a cointegrating relationship does not depend on the order of the assets. This property ensures that the identified stationary combination (the spread) is invariant to variable ordering, providing a consistent and theoretically robust framework for multi-asset pairs trading.

In this work, both the Engle–Granger and Johansen methods were applied to each asset pair. In most practical two-asset cases, both methods lead to the same qualitative conclusion about cointegration. However, if the residual spread is weakly stationary (close to unit root), one test may reject, and the other may not. In instances where the ADF and PP tests disagreed with the Johansen rank test, the latter was retained as the reference, given its system-based formulation and maximum-likelihood efficiency.

3.4 Normalization

Pairs trading begins with constructing a spread capturing the long-term linkage between two (or more) asset price levels and then normalizing that spread so that signals are comparable across time. Let s_t denote the spread, e.g., for an asset pair $s_t = p_{1,t} - \beta p_{2,t}$. Because the raw spread's scale and volatility drift over time, we convert it into a rolling z-score:

$$z(s)_{t,L} = \frac{s_t - \mu_{t|t-1,L}}{\sigma_{t|t-1,L}},$$

where:

$$\mu_{t|t-1,L} = \text{mean}(s_{t-L:t-1})$$
$$\sigma_{t|t-1,L} = \text{sd}(s_{t-L:t-1}),$$

whereas $\mu_{t|t-1,L}$ is the rolling mean over the rolling window L, i.e., $\mu_{t|t-1,L}$ = mean($s_{t-L:t-1}$); $\sigma_{t|t-1,L}$ is the rolling standard deviation over L, i.e., $\sigma_{t|t-1,L}$ =sd ($s_{t-L:t-1}$); L is the rolling window that in our work is set to 21, i.e., the number of trading days in a month. This rolling normalization provides a dimensionless measure of deviation from equilibrium (in standard deviations), is robust to changes in spread scale, and helps maintain stationarity of the trading signal even when the underlying prices drift over time.

When we compute the rolling mean/standard deviation over a 21-bar window that ends at t, we use only data up to

and including t, i.e., $\mu_{t,L}$ =mean($s_{t-L+1:t}$) and $\sigma_{t,L}$ =sd($s_{t-L+1:t}$). If we execute the trading operation at *t*+1 (i.e., next bar), this is implementable and not look-ahead, but it *does* include *"self-influence"* at *t* (which, however, is not an error, just a choice). When we include s_t in the window that defines μ_t and σ_t, the current observation partly pulls the mean toward itself and can potentially inflate the volatility estimate. That shrinks extreme z-scores a bit (fewer/smaller threshold breaches). It is not a look-ahead error; it is a smoother, less conservative modeling choice at time *t*, that slightly dampens z-scores. If a signal at time t is used to trade at *t*+1, both choices are implementable, but the clean, conservative convention should use $\mu_{t|t-1}$ and $\sigma_{t|t-1}$ to avoid self-influence. In summary, our procedure is:

- Conservative (i.e., stricter), causally clean, harder to look good in backtests;
- No look-ahead: we never use *t*+1 data;
- With $\mu_{t|t-1}$, $\sigma_{t|t-1}$ we compare today's spread s_t to a benchmark fixed before today's move. The decision threshold is independent of the very observation that triggers the trade and it is stricter and causally clean (i.e., no self-inclusion and avoids self-dampening).

In summary: Self-inclusion is less conservative because today's observation helps set its own benchmark (moves μ_t toward s_t and can inflate σ_t), yielding smaller z_t, fewer crossings, and artificially smoother backtests. In contrast, no-self-inclusion ($\mu_{t|t-1}$, $-\sigma_{t|t-1}$) is more conservative, and yields larger z_t on shocks, hence more/earlier signals. Despite being operationally more active, it is preferred for unbiased, high-quality backtests.

3.5 Trading strategy

The pairs-trading approach builds a market-neutral portfolio by monitoring the normalized price spread (z-score) between two or more cointegrated assets. At each trading day the raw price spread is computed, then standardized by subtracting a rolling mean and dividing by the rolling standard deviation over a chosen look-back window. This transformation produces a z-score with zero mean and unit variance, allowing consistent comparison across time. The strategy's underlying assumption is mean reversion: if the spread deviates significantly from its long-run mean, it is likely to revert. By working with z-scores rather than raw spreads, traders can implement objective thresholds (measured in standard deviations) for entering and exiting positions. This normalization also facilitates position sizing, since it shows how extreme a spread movement is relative to recent history.

Trading signals are triggered when the z-score crosses predetermined band thresholds. For a long spread trade, a signal is generated when the z-score drops below -2 (meaning the spread is more than two standard deviations below its mean) while the previous day's z-score was above -2; this indicates the first asset has become relatively undervalued and should be bought while the second is shorted. The long position is closed when the z-score rises above −1, signaling partial mean reversion. Conversely, a short spread trade is entered when the z-score rises above +2 while the prior day's z-score was below +2, and it is exited once the z-score falls below +1. This stepwise exit avoids premature closure on small fluctuations and ensures the spread has meaningfully retraced. The strategy maintains the previous position in the absence of a new signal, leading to a simple state machine where positions can only transition at these thresholds. Such symmetrical entry/exit bands help control risk and reduce whipsaw trades; some practitioners set wider bands for entry than exit to capture more pronounced divergences and exit earlier to lock in profits [14][15][16].

The signal rules compare the previous day's z-score with the current z-score to determine whether a crossing occurred. The signal produced on day t therefore depends on data up to day t (inclusive) but is applied only from t+1 onward; no trade is executed until the following day. This one-period shift eliminates the possibility that the strategy trades on information not available at the decision time. The positions are updated sequentially; they never revisit or adjust past positions based on future outcomes. Because the z-scores, signals and positions at time t+1 rely solely on the spread, mean, and standard deviation computed using information up to time t, the entire strategy is free from look-ahead bias.

A major strength of this framework is its simplicity and discipline: it quantifies mispricing using standardized units and prescribes clear rules for trade execution, promoting market neutrality and reducing behavioral bias. However, its success depends on parameter choices, such as the look-back window for the rolling mean/standard deviation and the ±2/±1 thresholds. Using a short window may lead to self-influence, where the current observation affects both the signal and the benchmark; this can shrink extreme z-scores and artificially inflate volatility, potentially dampening trading signals. A simple remedy is to compute the mean and standard deviation up to t-1 (i.e., lag the window) so the current z-score is measured against truly historical data. Traders should also be aware of structural breaks, i.e., if the long-run relationship between assets changes, the spread may stop mean-reverting and the z-score signals will fail. Finally, transaction costs and slippage can erode returns, especially if the strategy triggers frequent trades.

The strategy sets the current trading position to +1 when it receives an *enter-long* signal, to 0 when it receives an *exit-long* signal, to −1 on an *enter-short* signal, and back to 0 on an *exit-short* signal. If no new signal is present, the position remains unchanged, preserving the previous day's exposure. This explicit mapping ensures that each trade is only executed when the z-score crosses one of the predefined thresholds; otherwise the existing long or short position is maintained. Such a state machine is essential for systematic trading because it distinguishes between event-driven actions (enter or exit) and inert periods when the strategy simply holds its current position.

When analyzing this approach, it is important to recognize both its strengths and its sensitivities. By quantifying mispricing in standard-deviation units, the strategy imposes a disciplined, symmetric framework for entry and exit; it is market-neutral by design, since every long position is paired with a corresponding short. This makes the method attractive when markets are choppy or directionless. However, the rule-based nature also means that performance is dependent on

threshold calibration and the stability of the cointegrated relationship. Setting the entry threshold too close to the mean may generate whipsaw trades, while setting it too far reduces the number of signals and can lead to missed opportunities. The general literature cautions that z-score thresholds should be selected based on backtesting, often in the range of 1.5σ–2σ for entry and a smaller threshold for exits. Moreover, the success of the strategy presupposes that the spread is stationary and truly mean-reverting; if the underlying relationship breaks down, the signals become unreliable, and the strategy can incur losses.

4. Performance metrics

The pairs strategies are evaluated relative to simple *buy-and-hold* benchmarks, and three key metrics are used: annualized return, annualized standard deviation, and annualized Sharpe ratio. In a pair-trading context, these metrics help answer three fundamental questions: how much the strategy earns on average per year; how volatile or risky those returns are; and whether the returns adequately compensate for risk relative to a risk-free benchmark. By comparing the performance of the long-short spread against the individual asset returns, it is possible to assess whether cointegration-based trading strategies deliver better risk-adjusted performance than simply holding the individual assets.

Annualized return measures the compound growth rate of a trading strategy over the given period, scaled to a one-year horizon. It is computed as:

$$r_a = \left(\prod_{t=1}^{n}(1+r_t)\right)^{\frac{252}{n}} - 1,$$

where r_t is the daily return at period t. Using a geometric (compound) approach, rather than a simple arithmetic mean, accounts for the effects of volatility and compounding, producing the growth rate that an investor would have experienced over a year. This makes annualized return particularly useful when comparing strategies with different holding periods or uneven return profiles. To assess risk, we used the annualized standard deviation, defined as the daily standard deviation multiplied by the square root of the number of periods per year:

$$\sigma_a = \sigma_{daily}\sqrt{252}.$$

The annualized standard deviation shows how widely a strategy's returns fluctuate in a typical year. The annualized Sharpe ratio [4][5] then combines return and risk by dividing the strategy's annualized excess return (over the risk-free rate) by its annualized standard deviation, essentially expressing return per unit of volatility:

$$SR_a = \frac{r_a - r_{risk-free,a}}{\sigma_a}.$$

This metric enables straightforward comparison across different strategies and assets: a higher Sharpe ratio indicates better reward for the level of risk taken, while a negative or low Sharpe ratio suggests that risk-adjusted performance is poor.

Figure 6. EWA (AUS)-EWC (CAN) cumulative returns (training range)



5. Experimental results and case studies

5.1 AUS-CAN case study

Consider the single-pair spread AUS (EWA)-CAN (EWC). EWA and EWC are both single-country equity ETFs offered by iShares, giving investors targeted exposure to the stock markets of Australia and Canada respectively. EWA and EWC are widely viewed as a strong cointegrating pair because both represent commodity-driven economies with comparable market profiles. Each fund's portfolio is dominated by similar sectors, particularly financials and natural resource industries, reflecting how both Australia and Canada rely on exporting commodities like minerals and energy. Such shared economic and sectoral foundations lead to significant co-movements in their equity returns, confirmed by a historically high correlation between EWA and EWC (~0.8 as reported in Figure 3). Indeed, both countries' stock markets tend to respond in tandem to global commodity price cycles and external demand shifts. Empirically, this relationship manifests in a stable spread, and the Engle-Granger test confirms that a linear combination of EWA and EWC prices is stationary (as reported in Table 9). This indicates a potentially stable long-run equilibrium and a mean-reverting spread, ideal characteristics for a pairs-trading or statistical-arbitrage strategy.

Figure 6 displays the cumulative returns of the EWA and EWC ETFs from 2007 to 2017 (training window), showing co-movement and apparently a stable long-run relationship, consistent with cointegration. Although the series occasionally diverge during periods of volatility, such as the 2008 global financial crisis or commodity-driven downturns, those deviations tend to be temporary and are followed by reversion, suggesting a mean-reverting spread. The two cumulative return series maintain a roughly parallel structure across the full training range, with similar peaks and troughs, implying that their long-run equilibrium is preserved despite short-term fluctuations. This visual evidence of synchronization and bounded divergence is consistent with a cointegrated pair.

The spread in the training range is shown in Figure 7. The EWA–EWC spread over 2007–2017 displays classic features of a stationary series: it oscillates around a stable long-run

mean (red line) without an evident deterministic trend, and asymmetric phases (e.g., the 2012–2013 upswing) are followed by reversion toward that mean. The amplitude of deviations appears bounded, rarely breaching ±2 on the standardized scale, and volatility clustering, while present at times, is not persistent. This behavior is coherent with stationarity and with the interpretation that the spread is cointegrated (mean-reverting) and suitable for pairs trading. Visual evidence in Figure 7 is corroborated by unit-root tests on the spread (ADF and PP), which reject non-stationarity at conventional significance levels (i.e., AUS_CAN p-value < 0.05, as specified in Table 9).

Figure 7. EWA (AUS)-EWC (CAN) spread (training range)



The performance evaluation of the pairs trading strategy is based on the direct comparison of arithmetic daily returns in the testing range. First, returns are calculated for the individual ETFs (EWA and EWC denoted as AUS and CAN in Figure 8). The spread return, computed as the difference between the EWA return and the product of the hedge ratio (evaluated in the training window and propagated in the testing range) and the EWC return, is then multiplied by the position vector to yield the strategy return series rfint, which realizes P&L only when the position is non-zero (long or short). A second return series, rfintc, incorporates a simple transaction cost model: if a position change occurs (indicated by a signal and a change in position from the previous day), a cost of 0.001 (10 basis points) is subtracted from the return to reflect slippage or fees.

Figure 8. EWA-EWC pairs strategy comparison (testing range)

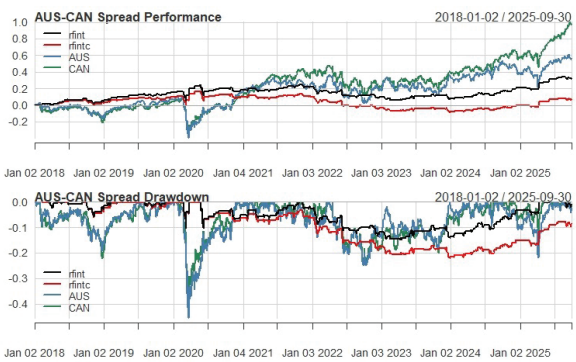


Figure 8 illustrates a distinct shift in performance dynamics between the pairs trading strategy and the buy-and-hold portfolio for EWA and EWC. During the initial phase of the testing window (2018 through mid-2020), the mean-reverting nature of the EWA–EWC spread allowed the pairs strategy to generate steady gains with lower volatility. In this period, both EWA and EWC experienced relatively flat or volatile movements without clear upward trends. These conditions are ideal for statistical arbitrage, as the spread exhibits repeated deviations from equilibrium and reverts predictably, enabling the strategy to capitalize on temporary short-lived spread deviations without being penalized by strong directional movements.

However, after the COVID-19 market correction in early 2020, both EWA and particularly EWC began trending upward, driven by a broader commodities recovery and a rally in global equities. In such directional markets, the pairs strategy, which inherently bets on relative value rather than outright momentum, underperforms. The spread stabilizes, offering fewer profitable mean-reversion opportunities, while buy-and-hold positions benefit directly from the persistent gains in the underlying assets. This market transition explains why, after mid-2020, the cumulative returns of both EWA and EWC begin to outperform the strategy returns, even before transaction costs are considered. The shift highlights a limitation of pairs trading: it succeeds in sideways or oscillating markets but tends to underperform during trending conditions.

Table 3. AUS (EWA)-CAN (EWC) performance metrics

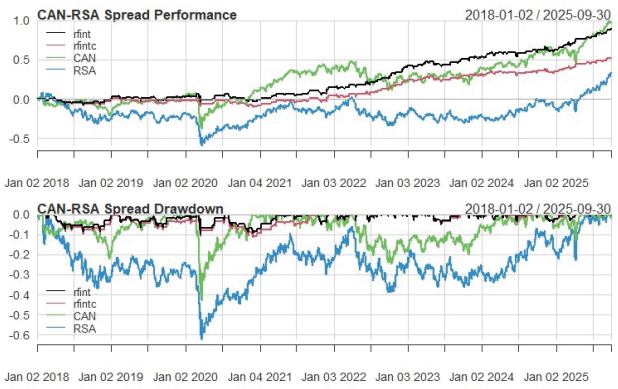
Statistic	rfint	rfintc	rfaus	rfcan
Annualized Return	0.038	0.010	0.061	0.093
Annualized Std Dev	0.076	0.075	0.242	0.201
Annualized Sharpe (Rf=0%)	0.506	0.131	0.251	0.464

The key performance metrics, i.e., the annualized return, annualized standard deviation, and Sharpe ratio for each series, are reported in Table 3. This combined analysis allows for a comprehensive evaluation of the strategy's effectiveness compared to simply holding either asset, and it highlights how trading costs can impact net performance even in a statistically robust strategy. The performance comparison reveals that the AUS–CAN pairs trading strategy outperformed the individual assets on a risk-adjusted basis, before accounting for transaction costs. The strategy without transaction costs (rfint) delivered a Sharpe ratio of 0.51, which is higher than that of EWA (rfaus) at 0.25 and slightly higher than EWC (rfcan) at 0.46. This indicates that the strategy was more efficient at converting volatility into return than simply holding either ETF. Moreover, while its raw annualized return (3.82%) is lower than the buy-and-hold return for both EWA (6.08%) and EWC (9.30%), the much lower volatility (7.55% vs. over 20% for the ETFs) underscores the effectiveness of the mean-reverting structure in managing downside risk. However, after including a transaction cost of 10 basis points per round trip, the strategy's performance deteriorates significantly. The annualized return drops to just 1%, and the Sharpe ratio falls to 0.13, indicating

that most of the statistical arbitrage gains are eroded by trading costs.

This evident contrast between gross and net performance highlights the importance of execution efficiency in implementing high-turnover strategies like pairs trading, and demonstrates how even modest transaction costs can erode statistical edge in high-turnover strategies. It also reinforces the need to balance statistical robustness with realistic cost modeling to assess the practical viability of such strategies. The drawdown panel in Figure 8 provides further insight into downside risk. The unadjusted strategy (rfint) experiences shallower and less frequent drawdowns compared to the ETFs, confirming its defensive profile. Notably, during the sharp market stress of early 2020, EWA and EWC suffered severe drawdowns exceeding −40%, while the unadjusted strategy's losses remained much more contained. In contrast, the cost-adjusted version (rfintc), while still stable, occasionally draws down more than the unadjusted line (rfint), likely due to returns being clipped at rebalancing points. Overall, the visual evidence supports the conclusion that the pairs strategy offers smoother performance and lower risk, if and only if trading costs are minimal or well-managed, confirming that cost efficiency is critical for the practical viability of this statistical arbitrage strategy.

Figure 9. EWC-EZA pairs strategy comparison (testing range)



5.2 CAN-RSA case study

The CAN (EWC)−RSA (EZA) pairs trading strategy demonstrates strong and consistent outperformance over the testing period, as shown in both the cumulative return and drawdown plots in Figure 9. The strategy without transaction costs (rfint) generates steadily increasing returns with limited drawdowns, while even the cost-adjusted version (rfintc) outperforms both individual ETFs. In contrast, the RSA ETF (blue) exhibits substantial volatility and deep drawdowns, particularly from 2018 through 2022, whereas the CAN ETF (green) performs more robustly, especially after 2020. Despite CAN's stronger long-term upward trend, its return still comes with higher volatility than the spread strategy. In contrast, pairs trading capitalizes on the relative mispricing between the two assets, generating stable gains as the spread mean-reverts across time.

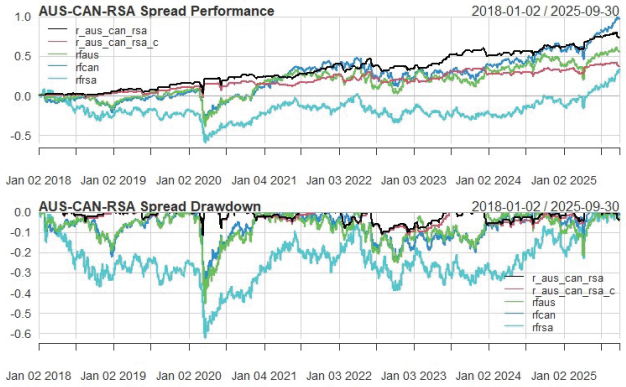
The performance metrics are reported in Table 4. The strategy's Sharpe ratio reaches 0.99 before costs and 0.65 after costs, significantly higher than CAN (0.46) and RSA (0.12). This strong risk-adjusted return reflects the statistical nature of the spread, which successfully isolates relative value opportunities between the two ETFs. The CAN−RSA spread seems to revert consistently to equilibrium, suggesting a robust long-run cointegration. Even during volatile markets, the spread's mean-reversion dynamics generate profitable entry and exit points. Therefore, the strategy's superior performance stems from its ability to avoid the directional exposure and volatility embedded in each ETF and instead exploits the stable co-movement patterns between them.

Table 4. CAN (EWC)-RSA (EZA) performance metrics

Statistic	rfint	rfintc	rfcan	rfrsa
Annualized Return	0.086	0.056	0.093	0.039
Annualized Std Dev	0.087	0.087	0.201	0.316
Annualized Sharpe (Rf=0%)	0.992	0.650	0.464	0.123

The strong performance of the CAN−RSA spread is underpinned not only by statistical mean-reversion, but also by macroeconomic parallels between the Canadian and South African economies. Both countries are highly resource-dependent, with significant exposure to commodities such as oil, gold, and base metals. The ETFs representing these markets, EWC for Canada and EZA for South Africa, reflect this structural similarity in their sector compositions, which are heavily weighted toward energy, mining, and materials. These shared economic drivers often lead to correlated responses to global commodity price movements, forming the basis for a potentially cointegrated relationship in their equity indices, and supporting the trading strategy's foundation: deviations in the relative pricing between EWC and EZA are likely temporary and driven by short-term noise rather than fundamental divergence. The strategy effectively profits from this short-lived mispricing by going long the undervalued ETF and short the overvalued one, and anticipating reversion. As seen in the cumulative return and drawdown profiles, this approach yields consistent gains and low volatility compared to the standalone ETFs, which remain exposed to individual country risks and broader macroeconomic shocks.

Figure 10. EWA-EWC-EZA multi-pairs strategy comparison (testing range)



5.3 AUS-CAN-RSA case study

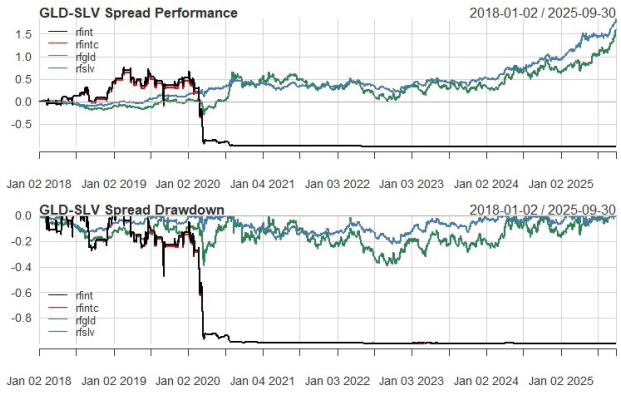
The performance of the AUS–CAN–RSA multi-asset spread illustrated in Figure 10 reveals steady and consistent gains over the testing period. The strategy without transaction costs (r_aus_can_rsa) delivers strong cumulative returns with limited drawdowns and low volatility, achieving an annualized Sharpe ratio of 0.81, as reported in Table 5.

Table 5. AUS-CAN-RSA performance metrics

Statistic	r_aus_can_rsa	r_aus_can_rsa_c	rfaus	rfcan	rfrsa
Annualized Return	0.075	0.043	0.061	0.093	0.039
Annualized Std Dev	0.093	0.093	0.242	0.201	0.316
Annualized Sharpe (Rf=0%)	0.805	0.460	0.251	0.464	0.123

Even after accounting for transaction costs (r_aus_can_rsa_c), the strategy maintains positive performance (Sharpe ratio of 0.46), comparable to the standalone Sharpe of the best-performing ETF (EWC at 0.46). Importantly, both strategy variants exhibit superior stability relative to the individual ETFs, especially EZA (South Africa), which suffers deep and extended drawdowns. This highlights the benefit of capturing a stationary, mean-reverting spread that arises from the relative dynamics among Australia, Canada, and South Africa. All these countries are major commodity exporters with strong exposure to natural resources, and their equity markets, represented by EWA (Australia), EWC (Canada), and EZA (South Africa), share common drivers like global demand for metals, oil, and energy. However, these economies respond differently to local shocks and policy decisions, leading to short-term deviations among their respective ETFs. These deviations are often temporary, allowing the strategy to exploit the reversion to equilibrium.

Figure 11. GLD-SLV pairs strategy comparison (testing range)



The relatively low volatility of the constructed spread and the consistent performance despite fluctuating ETF returns confirm that the three-asset spread effectively filters out idiosyncratic risk while preserving exposure to a stable common stochastic trend.

5.4 GLD-SLV case study

The GLD–SLV pairs trading strategy initially delivered strong performance in the testing range, as shown in the top panel of Figure 11. From 2018 through January 2020 the cumulative return of the strategy (both without –rfint– and with transaction costs –rfintc–) moved upward and outperformed the buy-and-hold performance of the single-leg GLD (rfgld) and SLV (rfslv) ETFs, as reported in Table 6. It is noteworthy, however, that in the training range the GLD–SLV spread failed the Engle–Granger residual-based stationarity test, as confirmed in Table 10. Under the standard cointegration framework, the two ETFs would therefore not qualify as a cointegrated pair and, strictly speaking, should have been excluded from the pairs-trading strategy.

Table 6. GLD-SLV performance metrics (2018 - January 2020)

Statistic	rfint	rfintc	rfgld	rfslv
Annualized Return	0.190	0.161	0.089	0.018
Annualized Std Dev	0.556	0.555	0.107	0.178
Annualized Sharpe (Rf=0%)	0.341	0.289	0.830	0.101

Academic research confirmed that long-run relationship between gold and silver started weakening after 1990 [18], and the study in [19] concluded that after 1990 gold and silver markets became increasingly separate, and that their prices were more likely determined by distinct fundamentals rather than common economic factors. When the two metals share phases of strong correlation and their prices tend to move together, it is because they both serve as stores of value, inflation hedges, and safe-haven assets [20].

In the testing subperiod from 2018 till January 2020, the gold–silver correlation was ~0.8 (which can be classified as high correlation) and the gold-silver ratio fluctuated in a relatively narrow band within its historical range, as illustrated in Figure 12. Although the ratio exhibits short-horizon reversals, it follows a mild upward drift, rather than reverting to a fixed mean. In this time window the spread did not pass the Engle-Granger stationarity test; the residual-based ADF test fails to reject the unit-root null at the conventional 5% level (only marginally at 10%), hence, in this range GLD and SLV cannot be regarded as cointegrated; any indication of mean reversion in the spread is, at best, weak and statistically fragile. Nevertheless, trades based on spread mean-reversion generated profits (see Table 6).

Figure 12. Gold-silver ratio



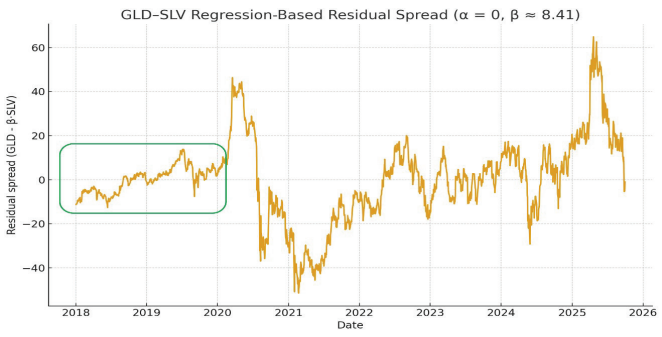
The market changed dramatically around March 2020. The onset of the COVID-19 pandemic drove an unprecedented flight to safe-haven assets and inflation hedges, causing gold prices to rally, while silver initially lagged and as an industrial commodity introduced volatility and retail speculation. The COVID panic peak (mid-March 2020) corresponds to a GLD-SLV ratio of roughly 12½. Mapping this ETF ratio to the classic bullion ratio (which was reported around 120:1 at the extreme), the GLD/SLV series captures the same scenario: gold massively outperformed silver during the COVID crash, pushing the ratio to a historic high. Moreover, Caporale and Gil-Alana (2023) [13] found that starting from January 2020, mean reversion in gold and silver price differentials is virtually absent. With the spread no longer mean-reverting after January 2020 as displayed in Figure 13, the pair strategy stopped generating profitable signals and what used to be short-term mispricing turned into prolonged trend deviations. By the time convergence occurred, it was often after long stretches of divergence, thus leading to the strategy's underperformance, as reported in Table 7. In practical terms, a trade that would "buy silver & sell gold" on an overshoot (or vice-versa) started to either not revert or take too long to revert, making profits elusive.

Table 7. GLD-SLV performance metrics (testing range)

Statistic	rfint	rfintc	rfgld	rfslv
Annualized Return	-0.593	-0.603	0.145	0.132
Annualized Std Dev	1.004	1.004	0.147	0.282
Annualized Sharpe (Rf=0%)	-0.591	-0.601	0.983	0.469

In summary, the GLD–SLV pairs trade lost its edge because the statistical foundation (a stationary spread) was undermined by extraordinary macroeconomic events and shifting market dynamics. What had been considered as a stable precious metals arbitrage turned into a much more unpredictable relationship after 2020, causing the strategy's early gains to plateau and rendering it largely ineffective in the subsequent years.

Figure 13. GLD-SLV residual spread (testing range)



5.5 NED-FRA case study

The EWN–EWQ pairs strategy (Dutch vs. French broad-market ETFs) was based on a statistically in-sample cointegrated spread (ADF p-value 0.034, PP p-value 0.01, as reported in Table 9), but in the testing period the strategy underperformed buy-and-hold, as illustrated in Figure 14. Notably, EWN (Netherlands – rfned) returned about 10% annualized vs. 7.4% for EWQ (France – rffra) as shown in Table 8. EWN also exhibited higher volatility and higher Sharpe (0.44 vs 0.34). This persistent outperformance of EWN implies a trending spread rather than a stationary mean: the Netherlands ETF had a stronger upward drift. Consequently, the mean-reverting trades often bet wrong (e.g., short EWN/long EWQ) when EWN kept rallying.

Table 8. NED-FRA performance metrics

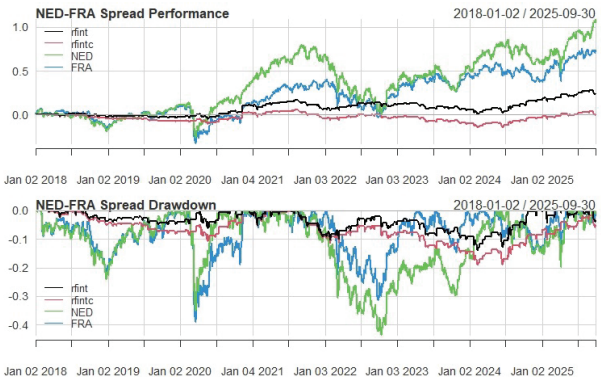
Statistic	rfint	rfintc	rfned	rffra
Annualized Return	0.0274	-0.0003	0.1002	0.0743
Annualized Std Dev	0.0682	0.0682	0.2249	0.2180
Annualized Sharpe (Rf=0%)	0.4025	-0.0037	0.4454	0.3407

The Dutch and French economies have faced different structural conditions since 2018, leading to economic and macro divergence. For example, the Netherlands (EWN) is heavily weighted toward global-tech and trade-sensitive stocks, whereas France (EWQ) leans more on luxury brands, energy, and domestic industries. ASML (semiconductors) alone is ~29% of EWN, while LVMH (luxury) is ~8% of EWQ. In other words, just because two series cointegrated in one period does not guarantee it holds later, and a temporary equilibrium broke down as markets evolved.

Pairs-trading research emphasizes that profitability is based on the stability of the cointegrating relationship. In our case, any slow drift or new trend (e.g. EWN consistently outpacing EWQ) would violate the stationarity assumption. Once the spread no longer reverts to its mean, mean-reversion signals generate losses. Furthermore, pairs trading is known to struggle during rapid moves; Gatev et al. [10] noted that pairs trading *"often fails to capitalize on rapid market movements due to its reliance on mean reversion."* The strategy's profitability is sensitive to entry/exit timing; delaying trades by one day substantially reduces returns, suggesting that a portion of profits may reflect microstructure effects or rapid market adjustments rather than slow mean reversion. In extreme volatility clusters, ETFs can swing together or apart for extended periods. For example, if EWN fell sharply in early 2020 (tech drawdown) and then rebounded on stimulus faster than EWQ, the spread would explode beyond the z-score entry thresholds. Even after any rebound, the expected quick snap-back might never occur in a timely way. The fact that EWN's drawdowns were deeper and volatility higher suggests it moved more erratically. These dynamics likely induced false signals or prolonged losing trades. In summary, the EWN–EWQ strategy underperformed because the out-of-sample environment violated its core assumptions. Structural or sectoral shifts and

volatility regimes likely broke the equilibrium that existed in the training range, turning the spread into a trending or unstable series. This analysis highlights a key lesson: cointegration-based strategies can be fragile. They require stable economic linkages and must withstand macro changes. Recent research confirms that adaptive risk controls and careful pair selection are vital [21][22].

Figure 14. NED (EWN)-FRA (EWQ) pairs strategy comparison (testing range)



5.6 Cointegration assessment

The Engle–Granger (EG) test results are reported in Table 9 and Table 10 for all single-pair spreads considered in this work that are listed in Table 11. For the pairs reported in Table 10 the EG residual-based unit-root ADF/PP tests fail to reject the null hypothesis of no cointegration at the chosen significance level (i.e., 5%), implying the regression residuals are non-stationary and the corresponding pairs are not cointegrated. It is worth noting that most of the spreads in Table 10 are commodity-related, for which structural breaks and non-stationary behavior are common, making residual stationarity harder to obtain in simple EG tests.

Table 9. Engle-Granger test results (training range) — single-pair spreads (cluster #1)

Test	AUS_CAN	CAN_RSA	AUS_RSA	GER_FRA	NED_FRA	GER_NED	AUS_BHP
ADF p-value	0.01	0.02	0.14	0.09	0.03	0.26	0.27
PP p-value	0.01	0.01	0.03	0.06	0.01	0.26	0.66

Test	CAN_SU	AUS_GLD	CAN_WTI	EWZ_VALE	RSA_GLD	SPY_DIA	AAPL_MSFT
ADF p-value	0.23	0.62	0.36	0.06	0.47	0.99	0.61
PP p-value	0.05	0.71	0.40	0.02	0.46	0.98	0.64

Table 10. Engle-Granger test results (training range) — single-pair spreads (cluster #2)

Test	GLD_GDX	GLD_IEO	XLE_IEO	LQD_SPY	GDX_IAU	GLD_SLV	XLE_OIH
ADF p-value	0.61	0.55	0.17	0.89	0.56	0.38	0.09
PP p-value	0.52	0.63	0.07	0.91	0.48	0.42	0.08

Test	KO_PEP	XLF_KRE	XOM_CVX	WMT_COST	VTI_SPY	OIH_USO
ADF p-value	0.80	0.41	0.66	0.47	0.49	0.37
PP p-value	0.84	0.53	0.55	0.58	0.41	0.45

Commodity prices are driven by highly volatile and idiosyncratic fundamentals, which makes it difficult to find a stable long-run equilibrium between different commodities. Key factors include:

- Each commodity has unique economic drivers. Gold and silver, for instance, are both precious metals but with different roles: gold is heavily influenced by investments and central bank demand, whereas silver has large industrial uses and is costly to store. These differences cause their price paths to diverge over time. Indeed, analysts note that gold and silver's relationship is not consistently cointegrated, making a reliable mean-reverting spread generally not viable.
- Energy commodities exhibit storage costs and seasonality that can cause systematic drifts between related assets. An oil exchange-traded fund (like USO, tracking oil futures) may suffer roll costs and deviate from an oil equity index (like OIH), especially if oil companies hedge or face financing constraints. These idiosyncrasies prevent commodity pairs from maintaining a tight long-run spread.
- Unlike financially linked assets, two different commodities have no inherent arbitrage to enforce a stable price ratio. There is no mechanism ensuring, say, gold and silver prices revert to a fixed proportion, and without such a mechanism their spread can trend or break out for long periods (a sign of no stationarity cointegration).

In contrast, equity and index pairs often exhibit more stable relationships due to shared economic drivers and overlapping exposures. Equities within the same sector, or indices in the same market, tend to move together over the long run. In summary, the economic nature of commodities – prone to independent supply/demand swings, storage effects, and geopolitical shocks – means that pairs like GLD–SLV or OIH–USO lack a steady long-run equilibrium, yielding non-stationary spreads in practice. Equities and indices, by contrast, share common economic drivers and often maintain a more stable relative pricing, making their spreads more likely to be stationary over time.

We tested whether each three-asset candidate spread listed in Table 11 has a stable long-run relationship using Johansen's trace test on (adjusted close) price levels. The Johansen test results are reported in Table 12 and Table 13. For each three-asset candidate we estimated a VAR (vector autoregression) and selected the lag length K by means of information criteria [12], then ran a VECM (vector error-correction model) [7][17] without deterministic trend in the cointegration space. The tables report K and the trace statistics for the sequential nulls $r \le 2$, $r \le 1$, $r = 0$ (columns test_r2, test_r1, test_r0), alongside the 10%, 5%, and 1% critical values (columns crit10_, crit5_, crit1_). Reading rule: if test_r0 > crit5_r0, we reject "no cointegration" and conclude there is at least one cointegrating vector; if test_r1 > crit5_r1, we infer at least two cointegrating vectors, and so on. Cells highlighted in green indicate rejection of the null hypothesis (i.e., no cointegration) at the 5% significance level.

Table 11. Single- and multi-pair spreads

Pair Spread
AUS-CAN
CAN-RSA
AUS-RSA
GER-FRA
NED-FRA
GER-NED
AUS-BHP
CAN-SU
AUS-GLD
CAN-WTI
EWZ-VALE
RSA-GLD
SPY-DIA
AAPL-MSFT

Pair Spread
GLD-GDX
GLD-IEO
XLE-IEO
GDX-IAU
GLD-SLV
XLE-OIH
OIH-USO
LQD-SPY
KO-PEP
XLE-KRE
XOM-CVX
WMT-COST
VTI-SPY

Multiple Pair Spread
XLE-OIH-IEO
SPY-KO-PEP
XLE-XOM-CVX
SPY-WMT-KO
WMT-KO-PEP
OIH-USO-WTI
OIH-USO-XLE
AUS-CAN-RSA
GER-NED-FRA
CAN-SU-WTI
AUS-RSA-GLD
SPY-AAPL-MSFT

Table 12. Johansen test results (training range) — multi-asset spreads (cluster #1)

Spread	K	test_r2	crit10_r2	crit5_r2	crit1_r2	test_r1	crit10_r1	crit5_r1	crit1_r1	test_r0	crit10_r0	crit5_r0	crit1_r0
AUS_CAN_RSA	4	4.48	6.5	8.18	11.65	18.87	15.66	17.95	23.52	34.08	28.71	31.52	37.22
GER_NED_FRA	2	0.00	6.5	8.18	11.65	10.11	15.66	17.95	23.52	22.12	28.71	31.52	37.22
CAN_SU_WTI	9	2.49	6.5	8.18	11.65	10.47	15.66	17.95	23.52	25.34	28.71	31.52	37.22
AUS_RSA_GLD	3	3.66	6.5	8.18	11.65	9.48	15.66	17.95	23.52	33.83	28.71	31.52	37.22
SPY_AAPL_MSFT	2	1.71	6.5	8.18	11.65	8.36	15.66	17.95	23.52	15.15	28.71	31.52	37.22

Table 13. Johansen test results (training range) — multi-asset spreads (cluster #2)

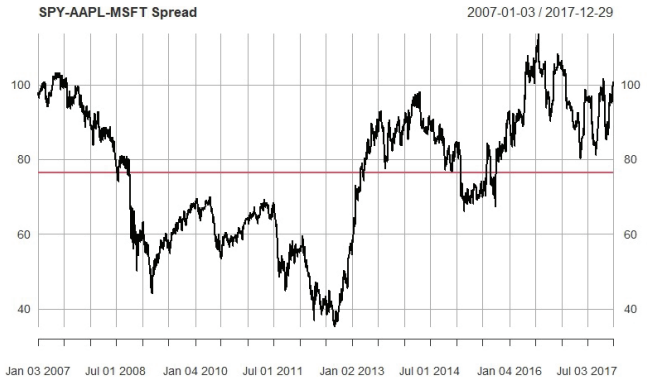
Spread	K	test_r2	crit10_r2	crit5_r2	crit1_r2	test_r1	crit10_r1	crit5_r1	crit1_r1	test_r0	crit10_r0	crit5_r0	crit1_r0
XLE_OIH_IEO	3	0.011	6.5	8.18	11.65	6.16	15.66	17.95	23.52	20.73	28.71	31.52	37.22
SPY_KO_PEP	2	0.097	6.5	8.18	11.65	9.60	15.66	17.95	23.52	26.77	28.71	31.52	37.22
XLE_XOM_CVX	3	0.003	6.5	8.18	11.65	7.33	15.66	17.95	23.52	15.53	28.71	31.52	37.22
SPY_WMT_KO	3	0.186	6.5	8.18	11.65	6.55	15.66	17.95	23.52	17.48	28.71	31.52	37.22
WMT_KO_PEP	3	0.005	6.5	8.18	11.65	5.63	15.66	17.95	23.52	13.59	28.71	31.52	37.22
OIH_USO_WTI	10	2.078	6.5	8.18	11.65	17.16	15.66	17.95	23.52	35.84	28.71	31.52	37.22
OIH_USO_XLE	3	0.000	6.5	8.18	11.65	9.18	15.66	17.95	23.52	37.47	28.71	31.52	37.22

When only the r = 0 trace (or max-eigenvalue) test statistic exceeds the 95% critical value, but the statistics for r = 1 (and higher) do not, the Johansen cointegration test indicates exactly one cointegrating vector. In other words, the cointegration rank is 1. In practical terms, the series are not all drifting apart independently and there is one linear combination that is stationary. Next, the test considers the null hypothesis: at most one cointegrating vector, versus the alternative of more than one. In this scenario, when the statistic for r = 1 does not exceed the 95% critical value, we fail to reject the null that there is at most one cointegrating vector. This means the evidence is insufficient to claim a second cointegrating relationship. In practical terms, the assets in the spread share one long-run equilibrium linear combination that is stationary (the cointegrating vector), while any additional independent combinations remain non-stationary. This outcome implies that the spread's cointegration rank is still 1.

In summary, only one cointegrating relationship is present in the system under the given Johansen test results. If both the r = 0 and r = 1 Johansen test statistics are greater than their 95% critical values, it means we reject the null hypothesis of no cointegration (r = 0), and also reject the null of at most one cointegration (r ≤ 1) at the 95% confidence level. In other words, the test finds significant evidence of more than one cointegrating relationship. This outcome implies that the cointegration rank is at least 2, i.e. there are two cointegrating vectors. Practically, the assets have two independent stationary (long-run equilibrium) linear combinations. Next step is testing r = 2 vs r > 2 to confirm if there are exactly two cointegrating vectors. If the r = 2 test statistic is not significant at 95% (i.e.,

the null hypothesis is not rejected), it suggests exactly two cointegrating vectors. Rejecting r = 0 and r = 1 but not r = 2 is the classic indication of two cointegrating vectors.

Figure 15. SPY-AAPL-MSFT spread (training range)



5.7 Outperformance without cointegration: SPY-AAPL-MSFT case study

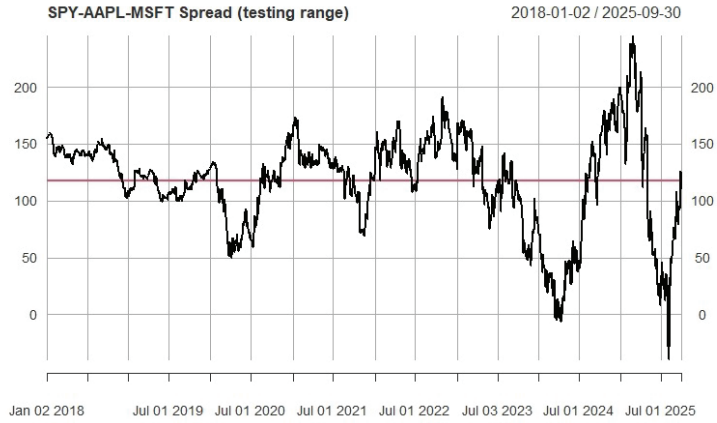
The SPY-AAPL-MSFT spread in the training range is shown in Figure 15. The spread does not appear mean-reverting, as:

- It shows multi-year trends and level shifts: a long, persistent decline from ~2007 to ~2012, then a multi-year climb to a new, higher plateau (~2013–2017). This pattern is indicative of a unit-root I(1) process with drift, not a stationary I(0) process fluctuating around a fixed mean.
- The horizontal reference mean (red line) is not a stable attractor: The series stays on one side of its mean for years, with rare mean crossings and persistent deviations, with a half-life far longer than in standard mean-reverting spreads.
- Variance is time-varying (heteroskedastic): quiet around the troughs, then larger swings near the peaks, inconsistent with a stationary I(0) process with constant variance.

Moreover, the Johansen test found no cointegration among SPY, AAPL, and MSFT in the training period, as confirmed by the Table 12 results.

The SPY-AAPL-MSFT spread is not mean-reverting in the testing range either (Figure 16). Although it does show a few reversals toward the horizontal reference level, its behavior is not tightly mean-reverting. From 2018 through 2022, the spread fluctuates around the reference line with crossings and recoveries after deviations, which is visually consistent with short-horizon convergence. However, from 2023 onward the deviations become larger and more persistent: the spread falls toward very low levels in late 2023, rises sharply through 2024, reaches a major peak above 200, and then collapses again in 2025 before partially recovering. These wide swings suggest that the spread is affected by changing relative trends among SPY, AAPL, and MSFT rather than by a stable equilibrium alone. Therefore, visual inspection provides only weak support for mean reversion. The plot suggests that the spread has local mean-reverting episodes, because large deviations are often followed by reversals, but it does not confirm a stable stationary process around a fixed mean.

Figure 16. SPY-AAPL-MSFT spread (testing range)

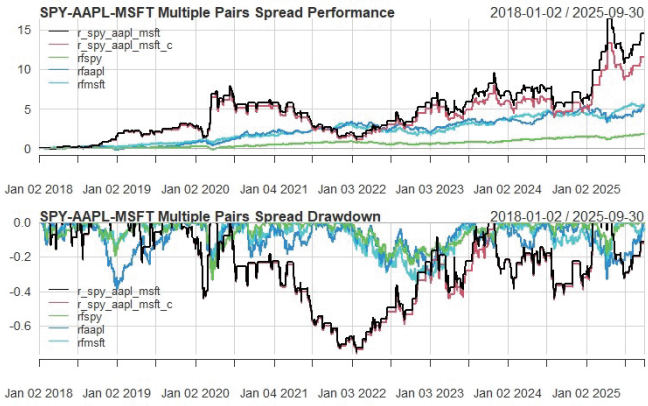


The volatility of the SPY–AAPL–MSFT spread is clearly time-varying, not constant. In the first part of the testing window, roughly 2018–2020, the spread fluctuates in a relatively narrower band around the reference level, with moderate excursions and several returns toward the mean. From 2021 onward, however, the amplitude of the fluctuations increases significantly, and the spread shows wider swings, sharper reversals, and larger deviations from the reference line, especially during 2023–2025. This behavior is consistent with volatility clustering, a well-known property of financial time series, suggesting that the spread may still offer short-horizon trading opportunities, but its risk is regime-dependent. Visual inspection indicates that the SPY–AAPL–MSFT spread exhibits local mean-reverting behavior rather than a pure I(1) random walk. However, the large and persistent deviations observed after 2023, together with time-varying volatility, prevent a visual classification as a stable stationary I(0) process. The spread captures the relative performance of two dominant technology mega-cap stocks against the broader S&P 500. Because Apple and Microsoft are themselves major constituents of SPY, the spread is anchored by a common equity-market component; however, deviations arise when the mega-cap technology factor outperforms or underperforms the diversified index. The large post-2023 excursions are consistent with the persistent leadership and rising index concentration of mega-cap technology stocks. Therefore, the spread dynamics are driven by both short-horizon convergence and structural changes in market leadership.

Then, how did the SPY–AAPL–MSFT spread trade manage to outperform the individual buy-and-hold returns, despite the lack of cointegration? The SPY–AAPL–MSFT spread outperformed the individual buy-and-hold benchmarks despite the absence of formal cointegration because the strategy exploited short-horizon relative-performance reversals rather than a stable long-run cointegrating equilibrium. Apple and Microsoft as major constituents of SPY share common equity-market and mega-cap technology drivers with the index, but their relative performance versus the broader market varies

over time. During the testing period, the dominance of mega-cap technology generated large spread excursions, while periodic rotations between technology leadership and broader-market participation produced enough partial convergence for the z-score trading rule to be profitable. The trading rule took short exposure to the relatively expensive leg(s) and long exposure to the relatively inexpensive leg(s), realizing gains when the spread partially converged. A buy-and-hold allocation in SPY, AAPL, or MSFT captures the directional return of the selected asset, but it cannot monetize time-varying relative dislocations among the three instruments. By contrast, the spread strategy updates exposure only when predefined entry and exit thresholds are reached, allowing it to capture recurrent mean-reversion episodes that static allocations do not exploit. Conditional on standardized spread deviations, measured through the z-score, the result should therefore be interpreted as evidence of profitable tactical convergence trading, not as evidence of a stable long-run cointegrating relationship. Therefore, these outcomes support the view that cointegration is informative but not strictly necessary for tradable spread dynamics, provided that economically linked assets display recurrent short-horizon reversals and sufficient dispersion to overcome transaction costs. The cumulative effect was superior to simply holding any one of the assets, as illustrated in Figure 17.

Figure 17. SPY-AAPL-MSFT multi-asset strategy comparison in the testing range



Over the 2018–2025 testing window, the SPY–AAPL–MSFT spread delivered higher cumulative returns than buy-and-hold in SPY, AAPL, or MSFT, both gross (r_spy_aapl_msft) and net of costs (r_spy_aapl_msft_c), as reported in Table 14. This aligns with the charts in Figure 17: despite a deeper drawdown in 2021–mid-2023, when the relative trend ran against a mean-reversion rule, the strategy recovered sharply and outperformed from 2023 onward as repeated divergence–convergence episodes emerged. In short, the spread captures relative-value convergence (mean-reversion) that buy-and-hold cannot realize; transaction costs attenuate but do not eliminate the advantage.

Table 14. SPY-AAPL-MSFT performance metrics

Statistic	r_spy_aapl_msft	r_spy_aapl_msft_c	rfspy	rfaapl	rfmsft
Annualized Return	0.4257	0.3879	0.1426	0.2691	0.2751
Annualized Std Dev	0.5854	0.5850	0.1964	0.3113	0.2860
Annualized Sharpe (Rf=0%)	0.7272	0.6631	0.7258	0.8646	0.9621

The SPY–AAPL–MSFT case illustrates that out-of-sample profitability can arise even when in-sample cointegration is absent. Short-horizon mean-reversion rules can exploit transient, regime-dependent corrections in relative prices without requiring a perfectly stationary spread. However, the lack of cointegration implies drift risk (non-stationarity), so strict risk controls are essential. The observed performance is consistent with markets periodically correcting relative mispricing among fundamentally related securities; the spread framework converts these divergence–convergence episodes into returns that buy-and-hold positions in SPY, AAPL, or MSFT do not capture.

5.8 Out-of-sample performance summary: spreads vs buy-and-hold

The out-of-sample performance of all single-pair and multi-asset spreads over the 2018 – September 2025 testing window is reported in Table 15 and Table 16 respectively. For each spread we tabulate annualized return, annualized volatility, and annualized Sharpe ratio computed from daily arithmetic returns. "No comm", i.e., no commissions, denotes gross results; "Comm", i.e., with commissions, denotes net results after the transaction-cost model used in this study. For comparison, the same metrics are shown for the buy-and-hold legs (Asset #1, Asset #2, ...). Our primary evaluation metric is the net Sharpe; outperformance is assessed relative to the best constituent leg. Highlighted in "green" are all the positive annualized returns and Sharpe. The tables show "*at a glance*" where mean-reversion trading adds value after costs, where costs erode the edge, and how spreads compare with their benchmarks in the testing period.

Table 15. Strategy performance (single-pair spreads)

	Annualized Return				Annualized Std Dev				Annualized Sharpe (Rf = 0%)			
Spread	No comm	Comm	Asset #1	Asset #2	No comm	Comm	Asset #1	Asset #2	No comm	Comm	Asset #1	Asset #2
AUS-CAN	0.038	0.010	0.061	0.093	0.076	0.075	0.242	0.201	0.506	0.131	0.251	0.464
CAN-RSA	0.086	0.056	0.093	0.039	0.087	0.087	0.201	0.316	0.992	0.650	0.464	0.123
AUS-RSA	0.022	-0.004	0.061	0.039	0.106	0.106	0.242	0.316	0.207	-0.038	0.251	0.123
GER-FRA	0.018	-0.010	0.056	0.074	0.049	0.049	0.223	0.218	0.380	-0.209	0.251	0.341
NED-FRA	0.027	0.000	0.100	0.074	0.068	0.068	0.225	0.218	0.403	-0.004	0.445	0.341
GER-NED	-0.010	-0.034	0.056	0.100	0.065	0.066	0.223	0.225	-0.158	-0.525	0.251	0.445
EWZ-VALE	-0.226	-0.245	0.019	0.059	0.791	0.791	0.352	0.406	-0.285	-0.310	0.053	0.146
SPY-DIA	0.012	-0.014	0.143	0.106	0.047	0.047	0.196	0.192	0.263	-0.297	0.726	0.550
AAPL-MSFT	0.105	0.074	0.269	0.275	0.147	0.147	0.311	0.286	0.713	0.505	0.865	0.962
GLD-GDX	-0.191	-0.213	0.145	0.175	0.653	0.653	0.147	0.363	-0.292	-0.325	0.983	0.482
GLD-SLV	-0.593	-0.603	0.145	0.132	1.004	1.004	0.147	0.282	-0.591	-0.601	0.983	0.469
KO-PEP	-0.035	-0.058	0.083	0.054	0.090	0.090	0.194	0.209	-0.393	-0.641	0.430	0.260
OIH-USO	-0.027	-0.051	-0.075	-0.034	0.412	0.412	0.457	0.397	-0.065	-0.124	-0.165	-0.086
VTI-SPY	-0.039	-0.066	0.136	0.143	0.062	0.062	0.200	0.196	-0.623	-1.061	0.681	0.726
XLE-IEO	0.025	-0.004	0.072	0.074	0.050	0.050	0.321	0.375	0.493	-0.073	0.224	0.198

Table 16. Strategy performance — (multi-asset spreads)

	Annualized Return					Annualized Std Dev					Annualized Sharpe (Rf = 0%)				
Spread	No comm	Comm	Asset #1	Asset #2	Asset #3	No comm	Comm	Asset #1	Asset #2	Asset #3	No comm	Comm	Asset #1	Asset #2	Asset #3
AUS-CAN-RSA	0.075	0.043	0.061	0.093	0.039	0.093	0.093	0.242	0.201	0.316	0.805	0.460	0.251	0.464	0.123
GER-NED-FRA	-0.015	-0.039	0.056	0.100	0.074	0.048	0.048	0.223	0.225	0.218	-0.306	-0.812	0.251	0.445	0.341
CAN-SU-WTI	-0.029	-0.053	0.093	0.055	-0.084	0.183	0.183	0.201	0.400	0.730	-0.159	-0.290	0.464	0.136	-0.115
AUS-RSA-GLD	0.062	0.035	0.061	0.039	0.145	0.102	0.101	0.242	0.316	0.147	0.610	0.341	0.251	0.123	0.983
SPY-AAPL-MSFT	0.426	0.388	0.143	0.269	0.275	0.585	0.585	0.196	0.311	0.286	0.727	0.663	0.726	0.865	0.962
XLE-OIH-IEO	0.028	-0.001	0.072	-0.075	0.074	0.046	0.046	0.321	0.457	0.375	0.609	-0.010	0.224	-0.165	0.198
SPY-KO-PEP	0.010	-0.014	0.143	0.083	0.054	0.275	0.275	0.196	0.194	0.209	0.037	-0.051	0.726	0.430	0.260
WMT-KO-PEP	-0.020	-0.042	0.179	0.083	0.054	0.136	0.136	0.222	0.194	0.209	-0.148	-0.305	0.804	0.430	0.260
OIH-USO-WTI	-0.055	-0.081	-0.075	-0.034	-0.084	0.201	0.201	0.457	0.397	0.730	-0.274	-0.403	-0.165	-0.086	-0.115
OIH-USO-XLE	-0.053	-0.077	-0.075	-0.034	0.072	0.113	0.113	0.457	0.397	0.321	-0.471	-0.685	-0.165	-0.086	0.224
XLE-XOM-CVX	-0.012	-0.037	0.072	0.086	0.071	0.057	0.057	0.321	0.305	0.317	-0.212	-0.639	0.224	0.283	0.225
SPY-WMT-KO	0.001	-0.022	0.143	0.179	0.083	0.097	0.097	0.196	0.222	0.194	0.010	-0.226	0.726	0.804	0.430

Out-of-sample results support a mixed but consistent picture. A few spreads deliver positive net Sharpe versus buy-and-hold, while others underperform after costs. Net performance weakens where assets trend for long periods or where term-structure effects dominate (several commodity-linked ETFs). Using tight entry/exit bands induces frequent position reversals and elevated turnover; the resulting transaction costs materially reduce net performance (Sharpe and return). Importantly, cointegration tests are informative but not decisive, some in-sample cointegrated spreads falter in testing, while non-cointegrated multi-asset spreads still earn short-horizon convergence gains. Overall, the results support a regime-aware, cost-controlled implementation with rolling diagnostics and periodic re-estimation, rather than full reliance on cointegration alone.

6. Conclusions

Although pairs trading and statistical arbitrage are well-established topics, most previous studies focus on earlier sample periods, narrow equity universes, two-asset spreads only, and often rely on weak out-of-sample protocols without properly considering transaction costs. The proliferation of liquid ETFs, the increased concentration of major equity indices in a few large stocks, and the widespread use of commodity-linked products have significantly changed the market structure relative to the settings examined in the classic literature. This motivated a re-examination of cointegration-based and mean-reversion strategies using a single, comprehensive and transparent framework applied consistently across equity, index, and commodity ETFs, and across a recent window that includes the global financial crisis aftermath, the COVID-19 shock, and the subsequent trending period. The present study was designed to fill this gap by combining Engle–Granger and Johansen diagnostics, explicit transaction-cost modeling, and a strict train/test split to assess whether the main concepts underlying pairs trading, e.g., stationarity, cointegration, and mean-reversion, still translate into robust performance under current market conditions and across a broader spectrum of assets.

This paper evaluated the conditions under which single-pair and multi-asset spreads obtained from liquid ETFs, equity indices, and commodities can generate reliable out-of-sample performance. We implemented a unified, reproducible methodology that: (i)

constructs spreads in the 2007–2017 training window via OLS on price levels with no constant term for two-asset pairs and Johansen eigenvectors for three-asset combinations; (ii) assesses long-run linkage using Engle–Granger and Johansen trace tests; and (iii) executes out-of-sample trades using an ex-ante 21-day rolling z-score computed without self-inclusion, asymmetric bands, state-persistent positions, and without look-ahead bias. Performance is measured from daily arithmetic returns and reported as annualized return, volatility, and Sharpe, both gross and net of an explicit transaction-cost model and benchmarked to the single-asset buy-and-hold legs.

The study shows that cointegration may not be sufficient, and not always necessary, for out-of-sample profitability in pairs and multi-asset trading. We analyzed cases in which spreads passed in-sample stationarity/cointegration tests yet failed to deliver excess performance in the testing window, consistent with instability of long-run relations under market regime changes. Conversely, we also observed non-cointegrated spreads in training that nonetheless produced out-of-sample mean-reversion gains, driven by transient, regime-dependent deviations in relative pricing rather than a stable stationary equilibrium.

A key finding is that while cointegration can be a diagnostic, it is neither a strict prerequisite nor a guarantee for mean-reversion strategies. In markets characterized by sector rotation, rising concentration in mega-caps, and shifting rate regimes, the structure is time-varying: a linear combination can display short-horizon mean reversion even when standard in-sample cointegration tests fail, and conversely an in-sample cointegrating relation can break out-of-sample due to parameter instability and structural breaks. Our framework with strict train/test split, ex-ante standardization with lagged rolling statistics (no self-inclusion), and explicit transaction-cost modeling, aims to measure actual out-of-sample performance and to avoid look-ahead and in-sample overfitting. In practice, pairs or multi-asset spreads that do not pass in-sample cointegration tests should not be excluded a priori when they show robust short-term trading signals, such as frequent and statistically consistent z-score excursions that revert within a practical holding period and deliver positive risk-adjusted returns net of costs. Conversely, even cointegrated spreads require continuous monitoring and periodic re-estimation, because parameters drift and rank changes. Moreover, after breakpoints, i.e., when the spread deviates from its long-run relation and the series correct the error by moving back toward equilibrium, the pull-back may become slower or statistically insignificant and ultimately erode performance.

Our results suggest that while cointegration can guarantee a long-run equilibrium, short-term profitability can occasionally be extracted from non-stationary spreads. However, we need to be aware that trading non-cointegrated spreads can introduce "drift risk," where mean-reversion relies on volatility clustering rather than true economic factors. Hence, stationarity remains the preferred standard for risk management, even if non-stationary pairs can occasionally outperform due to momentum effects. A regime-aware, risk-controlled implementation, with continuous diagnostics, periodic re-estimation, and explicit cost modeling, provides more reliable out-of-sample performance than relying on static cointegration rules alone.

Taken together, the results of this work show that pairs trading, when implemented with realistic transaction costs, rigorous out-of-sample testing and explicit attention to regime shifts, remains a viable and useful tool for both traders and investors. Rather than offering riskless arbitrage, the framework developed here characterizes pairs trading as a disciplined relative-value strategy. The empirical evidence, across equity, ETF and commodity-linked examples, suggests that such a structured approach can still deliver attractive risk–return profiles and meaningful diversification versus directional benchmarks.

Appendix - Stationarity vs non-stationarity

In time series analysis, $I(0)$ and $I(1)$ refer to the order of integration of a process, i.e., the number of times a series must be differenced to achieve stationarity.

- I(0): Stationary (Integrated of Order 0):
  A series is I(0) if it is stationary in levels, meaning its mean, variance, and autocorrelation structure do not change over time. Examples: returns of stocks, interest rate spreads, or residuals from a cointegrating relationship. Such series fluctuate around a constant mean and revert to it after shocks.
  Mathematically:
  $y_t \sim I(0)$ *if it is covariance-stationary.*
  Implications: The series has finite variance and a well-defined long-run mean. Shocks are temporary; the effect of a disturbance expires over time. Most statistical inference (OLS, t-tests, correlations) assumes variables are I(0).
- I(1): Non-stationary (Integrated of Order 1):
  A series is I(1) if it becomes stationary after first-order differencing:
  $\Delta y_t = y_t - y_{t-1} \sim I(0)$
  Examples: stock prices, exchange rates, or index levels. These series often show stochastic trends, i.e., they drift without mean reversion, and shocks have permanent effects.
  Implications: The series' mean and variance change over time. Correlations between I(1) series can be spurious (they appear correlated just because both trend upward). For meaningful relationships between I(1) series, we test cointegration, which looks for a stationary linear combination (I(0)) of I(1) variables.

In the context of pairs trading

- Asset prices are typically I(1): they trend and do not mean-revert.
- Returns are typically I(0): they fluctuate around zero and are stationary.

A cointegrating relationship exists if a linear combination of two (or more) I(1) price series forms an I(0) spread, i.e., their difference (after appropriate scaling) is stationary:

$$y_t = p_{1,t} - \beta p_{2,t} \sim I(0).$$

This stationary spread can then be exploited for mean-reversion-based trading.

References

[1] https://finance.yahoo.com/

[2] R: The R Project for Statistical Computing

[3] Zakamulin, V. (2014). The real-life performance of market timing with moving average and time-series momentum rules. Journal of Asset Management, 15(4), 261–278.

[4] Sharpe, W. F. (1966). Mutual Fund Performance. Journal of Business, 39(1), 119–138.

[5] Sharpe, W. F. (1994). The Sharpe Ratio. Journal of Portfolio Management, 21(1), 49-58.

[6] Engle, R. F., & Granger, C. W. J. (1987). Cointegration and Error Correction: Representation, Estimation, and Testing. Econometrica, 55(2), 251-276.

[7] Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica, 59(6), 1551-1580.

[8] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, 74(366), 427-431.

[9] Phillips, P. C. B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression. Biometrika, 75(2), 335–346.

[10] Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. Review of Financial Studies, 19(3), 797–827.

[11] Vidyamurthy, G. (2004). Pairs Trading: Quantitative Methods and Analysis. John Wiley & Sons, Hoboken, NJ, USA.

[12] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

[13] Caporale, G. M., & Gil-Alana, L. A. (2023). Gold and silver as safe havens: A fractional integration and cointegration analysis. PLoS ONE, 18(3), e0282631.

[14] Avellaneda, M., & Lee, J.-H. (2010). Statistical arbitrage in the US equities market. Quantitative Finance, 10(7), 761–782.

[15] Chan, E. P. (2013). Algorithmic Trading: Winning Strategies and Their Rationale. Hoboken, NJ: John Wiley & Sons.

[16] Caldeira, J. F., & Moura, G. V. (2013). Selection of a Portfolio of Pairs Based on Cointegration. SSRN Working Paper No. 2196391.

[17] Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control, 12(2–3), 231–254.

[18] Escribano, Á., & Granger, C. W. J. (1998). Investigating the relationship between gold and silver prices. Journal of Forecasting, 17(2), 81–107.

[19] Ciner, C. (2001). On the long run relationship between gold and silver prices: A note. Global Finance Journal, 12(2), 299–303.

[20] Bampinas, G., & Panagiotidis, T. (2015). Are gold and silver a hedge against inflation? A two-century perspective. International Review of Financial Analysis, 41, 267–276.

[21] Brunetti, M., & De Luca, R. (2023). Pre-selection in cointegration-based pairs trading. Statistical Methods & Applications, 32(5), 1695–1727.

[22] Chiu, M. C., & Wong, W. K. (2015). Dynamic cointegrated pairs trading: Mean–variance time-variant strategy. Journal of Economic Dynamics and Control, 55, 1-14.

Footnotes

1 Master of Financial Technical Analysis (International Federation of Technical Analysts)

2 Certified Financial Technician (International Federation of Technical Analysts)

3 Chartered Market Technician (CMT Association)

4 Certified SIAT (Società Italiana Analisi Tecnica) Technical Analyst

5 Convenience yield is the implicit or non-monetary benefit from holding the physical commodity rather than just holding a futures contract on it.

Predicting the Price

By David Linton, MSTA, MFTA

Reviewed by Regina Meani CFTe, MSTA, AMT

In *Predicting the Price,* Linton's aim is to look at *marrying a range of technical analysis techniques to increase the accuracy of price projections as well as gaining an understanding of when such predictions are failing.*[i]

In the introduction Linton explains how he developed his methods of using more than one technique at a time. Part One, comprising chapters one through eight are focused on an introduction to, and useful summary of, Technical Analysis.

In Chapter one the author gives us an insight into the historical development of technical Analysis which is an important stepping off point for the rest of Part One. He concludes with a summary of points relevant to Predicting the Price:

1. *A simple graph gives an understanding of the data*
2. *Dow Theory is central to technical analysis*
3. *Early charts were a shorthand for recording price and removed noise*
4. *Ideas of thrust and targets are nearly a century old*
5. *A chart represents the balance of buyer and sellers*[ii]

In Chapter Two we learn how timing is everything and that liquidity is vital. We also learn that finding your time horizon and lookback period are essential first steps in Predicting Price. The following chapters through to number eight also provide summary points which are pertinent to the topic and include the "Linton levels" for support and resistance and his midpoint averages. Linton points out that understanding chapters seven and eight on Point and Figure charts and the concept of vertical targets will help greatly in understanding the Linton Price Targets.

Chapter nine provides us with a check list of the things we must consider for understanding the Linton Price Targets and advises that they are completely rules based. Their great advantage is they can be used with most other Technical Analysis techniques.

We learn how to adapt Point and Figure charts to time based charts and how the method of construction can impact on our price targets. Linton argues the benefits of arithmetic to log scaling and points to the three notable differences, with the most important being the exaggeration generated on the log scale.

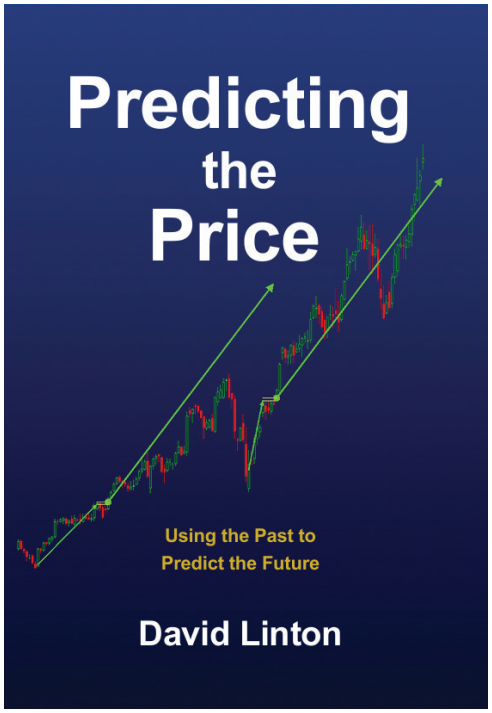
Linton promotes that *the key thing to understand about the price targets is that they are a function of buying or selling thrust.* [iii] Chapter ten provides us with the key factors and formulae behind the Linton Price Targets. The author stresses that the Linton Price targets adhere to the same strict rules as the Point & Figure vertical count. The author admits that one of the hardest challenges was projecting the Linton Price targets to a point in time. Linton Price Target *levels are about chart unit size, construction and target factor. The time projection or slope for a price target comes down to methodology.*[iv] There are five methods explored with the 'Price to time ratio" the simplest and most preferred method, working on the ratio of the distance to the target height and the distance in time.

A key part of price target analysis is looking for signs that a target is failing or exhausting and the author provides an identification system using the Linton method along with traditional forms of Technical Analysis. This leads to one of the most valuable features of the Linton Price targets is that it becomes more powerful when combined with other time-based methods.

David Linton has re-imagined a technique (developed over 80 years ago) by resolving some of its mystique to present us with a potent tool to add to our Technical Analysis arsenal.

Notes

[i] Linton,D, *Predicting the Price*, p6
[ii] Ibid., p27
[iii] Ibid., p134
[iv] Ibid., p163



Author Profiles

Sarah Alyasiri, MFTA, CFTe



Sarah Alyasiri, MFTA, CFTe, is a Financial Markets Strategist specializing in market analysis, investor education, and financial media. She regularly contributes market analysis through television interviews, seminars, webinars, and educational content, bridging market developments with practical perspectives for investors and professionals. Passionate about advancing financial literacy, she is committed to making financial markets more accessible through education, research, and clear communication.

David Linton MSTA, MFTA



David Linton MSTA, MFTA is founder of Updata and one of the UK's leading Technical Analysts. He holds the Master Financial Technical Analyst designation awarded by IFTA. He is author of 'Cloud Charts – Trading success with the Ichimoku technique' and teaches the Ichimoku module for the UK Society of Technical Analysts diploma. He is a member of the American Association of Professional Technical Analysts. David is the inventor of Linton Price Targets which are now available on several charting platforms including the Bloomberg Professional Terminal and TradingView.

Davide Pandini, PhD, CMT, MFTA, CFTe, CSTA



Davide Pandini holds a master's degree in Electronics Engineering (summa cum laude) from the University of Bologna (Italy) and a PhD in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh. In 1995, he joined STMicroelectronics in Agrate Brianza (Italy), a leading European semiconductor Integrated Device Manufacturer (IDM) company, where he is a Technical Director and Fellow of STMicroelectronics Technical Staff. Since June 2015, he has been the chairman of the ST Italy Technical Staff.

Dr. Pandini has earned the CMT, MFTA and CFTe designations, and is a professional member (CSTA) of SIAT (Società Italiana di Analisi Tecnica), a member society of the International Federation of Technical Analysts (IFTA), a global organization of market analysis societies and associations. In 2021, he received the XII SIAT Technical Analyst Award in the Open category, and the IFTA John Brooks Memorial Award. He was a speaker at the IFTA 2023 and 2024 annual conferences, the Bogu International Investment Forum 2024 in China, and the Open-Source Quantitative Finance 2025 conference in Chicago. His work has been published in the Journal of Technical Analysis and the IFTA Journal. He serves on the SIAT Scientific Committee and on the IFTA Board of Directors.

Patrick Plog, MScFE, MBA, MFTA, CFTe



Patrick Plog, MScFE, MBA, MFTA, CFTe, is a portfolio manager and quantitative researcher with more than 20 years of experience in financial and energy markets. His professional background spans institutional derivatives and European equity trading at HSBC in Düsseldorf, systematic portfolio management, and quantitative market research and proprietary trading. He currently manages power and gas portfolios in Germany and develops data-driven tools for market analysis, forecasting, and portfolio decision support. His research focuses on non-stationary time series, regime and latent-state modelling, forecasting, anomaly detection, market microstructure, and machine learning, with particular interest in translating quantitative methods into practical portfolio decisions. Patrick holds an MSc in Financial Engineering from WorldQuant University and an MBA from Liverpool John Moores University.

Maha Saeed, MFTA, CETA, CFTe, MSC



Maha Saeed, MFTA, CFTe, CETA, is a technical analyst based in Cairo, Egypt, specializing in technical analysis and portfolio management across Arab, U.S., and global financial markets. She is the CEO and Regional Director of the Arab Trading Academy and a lecturer at the Arab Academy for Management, Banking and Finance in Jordan, and Noor Business Academy. Maha holds a Master of Science degree, with her research focusing on integrating artificial intelligence with market cycle analysis to enhance investment decision-making in emerging markets and to compare traditional and non-traditional indices. She chairs the Research Committee at the Egyptian Society of Technical Analysts (ESTA) and is also a member of its Education Committee.

Tributes

In Memoriam

Over the decades, the field of technical analysis has been shaped by dedicated professionals whose research, discipline, and market insights left a lasting mark on the financial industry. These analysts studied price patterns, market trends, and investor behavior to develop tools and strategies that continue to influence traders and investors worldwide.

Although many of these respected figures have passed away, their contributions live on through the methodologies they pioneered, the educational materials they produced, and the generations of market participants they inspired. The individuals linked below, both of whom are now deceased, represent important contributors to the evolution and advancement of technical analysis. Their legacy remains an enduring foundation of modern technical analysis.



Constance "Connie" M. Brown

May 4, 1953 — October 15, 2025

Constance "Connie" M. Brown, of Tryon, North Carolina, passed away on October 15, 2025.

Connie was a highly respected figure in the global technical analysis community and a valued contributor to the International Federation of Technical Analysts (IFTA). She served on its Board of Directors and played an active role on the Education Committee. Her dedication to advancing technical analysis education—along with her thought leadership and publications—left a lasting impact on practitioners worldwide.

She held numerous professional designations, including the Certified Financial Technician (CFTe) and the Master of Financial Technical Analysis (MFTA) from IFTA, as well as the Chartered Market Technician (CMT).

Connie began her career as a manager for professional color films at Eastman Kodak. In 1988, she left Kodak to pursue her passion for financial markets. She became a professional institutional trader specializing in S&P 500 trading in New York City and Boston, and later served as a global equity futures hedge fund manager for six years. In 1996, she founded Aerodynamic Investments, Inc. to advance the field of market cycle timing.

A passionate educator, Connie traveled worldwide conducting seminars and authored influential works in technical analysis. Her books include All About Technical Analysis: The Easy Way to Get Started (2002); Fibonacci Analysis, which received a Gold Medal in the 2009 Axiom Business Book Awards; Technical Analysis for the Trading Professional; Breakthroughs in Technical Analysis, to which she contributed as editor; Mastering Elliott Wave Principle (2012); and her final work, The Thirty-Second Jewel (2019), widely regarded as a primary resource on W.D. Gann's methods, formulas, and time factors. In recognition of her lasting impact on the field, she was inducted into the Hall of Fame of the Canadian Society of Technical Analysis in 2018.

On November 16, 2021, Connie was honored by the Town of Tryon and Mayor J. Alan Peoples for her many accomplishments and her significant contributions to the study and advancement of financial technical analysis.

Before her career in finance, Connie was a world-class swimmer. She maintained a lifelong love of horses and beagles, especially her beloved companions, Sam and Bailey. She also enjoyed fencing and practiced Taekwondo.

Connie will be deeply missed within the technical analysis community, yet her legacy of knowledge, professionalism, and mentorship will remain a guiding light for analysts across generations.

May she rest in peace.



2019 IFTA Conference, Cairo, Egypt



2014 IFTA Conference, Kuala Lumpur, Malaysia



John J. Murphy

August 7, 1942 — February 7, 2026

The global technical analysis community pauses with heavy hearts.

We mourn the passing of John J. Murphy (August 7, 1942) — a true pioneer, mentor, and guiding light in the world of market analysis.

For decades, John's work shaped the way we understand charts, trends, and the powerful relationships between markets. His landmark books, including Technical Analysis of the Financial Markets and Intermarket Analysis, became more than references — they became foundational texts for generations of traders and analysts across the globe. Many of us did not just read his words; we built our careers upon them.

Beyond his titles and achievements, John was a teacher at heart. Through his years at Merrill Lynch, his educational work with the New York Institute of Finance, and his long-standing role at StockCharts.com, he made complex ideas accessible, practical, and deeply human. He reminded us that technical analysis is not merely about lines and indicators — it is about discipline, perspective, and respect for the market.

Many of us learned our first lessons in trend analysis, moving averages, and intermarket relationships through John's voice. His clarity gave confidence. His structure gave direction. His humility made him relatable.

We honour not only his professional legacy, but the generosity of spirit with which he shared his knowledge. His influence lives on in classrooms, trading desks, research reports, and in every analyst who approaches the markets with curiosity and integrity.

To his family, friends, colleagues, and the countless students he inspired around the world — we extend our deepest condolences.

Rest peacefully, John.

Your charts may no longer update, but your wisdom will continue to guide generations to come.



John Murphy speaking at the 2014 IFTA Conference in London

Top Right: Murphy with Dr. Rolf Wetzer, IFTA President (2013–2015) and Deborah Owen, Conference Chair

Bottom Right: Murphy with Indrawijaya Rangkuti, who shared, "He taught me about intermarket analysis."

IFTA Staff

Executive Director
Linda Bernetich, CAE

Production Manager (Interim)
Melissa Graham, MBA

Managing Editor (Interim)
Melissa Graham, MBA

Director of Accounting Services
Dawn Rosenfeld

Website Manager
Emma Richmond

IFTA Board of Directors

IFTA HEADQUARTERS
International Federation of Technical Analysts
1300 Piccard Drive, Suite LL 14
Rockville, MD 20850 USA
Phone: +1 (240) 404-6508
Fax: +1 (301) 990-9771
Email: admin@ifta.org | Web: www.ifta.org

President
Wieland Arlt, CFTe (VTAD)
wieland.arlt@ifta.org

Treasurer and Secretary
Oliver Reiss, Ph.D., MFTA, CFTe (VTAD)
oliver-reiss@freenet.de

Vice President Europe, Website Director
David Watts, BSc (Hons) CEng MICE MIWEM, FSTA (STA)
Dwatts360@gmail.com

Vice President Asia-Pacific
Akihiro Niimi, MFTA, CFTe (NTAA)
akihiro_niimi@ntaa.or.jp

Vice President The Americas, Webinar Director
Zoë Bollinger,CFP® (TSAASF)
zoe@bollingercapital.com

Vice President Middle East and Africa
Ron William, CFTe, MSTA, CFTe (SAMT)
roniwilliam@gmail.com

Education Director
Saleh Nasser, CMT, CFTe (ESTA)
snasser72@gmail.com

Examination Director
Gregor Bauer, Ph.D., CFTe, (VTAD)
gregor.bauer@vtad.de

Membership Director
Indrawijaya Rangkuti, MBA, CTA (AATI)
indra.senna@gmail.com

Marketing Director
Eddie Tofpik, MSTA (STA)
Eddie.Tofpik@admisi.com

Quant Director
Davide Pandini, Ph.D., MFTA, CFTe, CMT (SIAT)
davide.pandini@hotmail.com

Non-Institutional Director
Shinji Okada, CMTA, CFTe, MFTA (NTAA)
shinjiokadaresearch@gmail.com

Digital Media Director
Anisah Ozleen Othman, MSTA, CFTe (MATA)
remisieranis@gmail.com