

INTEGRATION OF THE REGIME-SWITCHING
FRAMEWORK FOR PORTFOLIO CONSTRUCTION
WITH COMMODITY FUTURES

ELLIE BAE

ADVISOR: PROFESSOR JOHN M. MULVEY

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN ENGINEERING

DEPARTMENT OF OPERATIONS RESEARCH AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY

APRIL 2024

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Ellie Bae

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Ellie Bae

Abstract

Commodity futures have long been considered excellent portfolio diversifiers due to their low correlation with other traditional financial assets. However, given the susceptibility of commodity futures markets to upheavals from both external (e.g., geopolitical events) and internal factors (e.g., shifts in supply and demand) — with volatility accentuated by global crises such as the Covid-19 pandemic and the Russian invasion of Ukraine — it is pivotal that investors capture these market dynamics accurately. In this study, we explore the application of the regime-switching framework to construct portfolios including commodity futures. Our research employs advanced methodologies to precisely identify and predict periods of price contagion within the commodity futures markets. Using the regime-switching framework, we optimize strategies of mean reversion, trend-following, and risk parity. We then assess these strategies against traditional portfolios, evaluating them through various performance metrics and varying degrees of diversification. Our findings reveal that the regime-switching framework significantly enhances strategy performance, offering higher risk-adjusted returns and superior drawdown management, although the efficacy of our approach appears to be influenced by the portfolio's composition, the selected commodity index, and the investor's risk tolerance. This paper thus demonstrates the regime-switching framework's potential to provide investors with a novel methodology for leveraging the unique benefits of commodity futures in volatile market conditions.

Acknowledgements

First, I would like to thank my advisor, Professor John Mulvey. All of your insights and advice have been immensely helpful as I navigated the research process. Thank you for your guidance for the past two years as I worked on my Junior Independent Project and Senior Thesis, and for the encouragement to explore novel ideas and topics. I also would like to express my appreciation for Rajiv for the support and feedback, as well as the ORFE department for all of the courses and resources that prepared me to write this thesis.

To My Manna Family: Moving from Korea for college was definitely challenging, but Princeton felt like home thanks to you all. Thank you for being such a precious community and family for me.

To my friends back in Korea: Even though we are more than 10,000 kilometers apart, I still feel so close to you all. Every time I go back home, I fully enjoy the time I spend with you all to the point of forgetting that we haven't seen each other for months.

To my roommates Joy, Katherine and Katie: I am so thankful for the 4+ years of friendship and precious memories that I share with you. This thesis wouldn't have been possible without the late-night studies in the common room, the Small World Coffee breaks, as well as all the laughters that we shared. I will definitely miss all of us living together a lot after we graduate, but I also look forward to us seeing each other grow in different ways as we enter a new stage of life. Joy, thank you also for thoroughly proofreading my entire thesis and giving me feedback, even despite your busy schedule.

To Sunny: I am grateful for you, and for the love and encouragement I have received from you this year. You are talented in many things, but exceptionally in making me laugh. Thank you for being kind and patient towards me.

To Mom and Dad: I want to express my love and gratitude for the unconditional love and support you have showered me with. I wouldn't have become who I am right now without all your love and wisdom. This opportunity to study at Princeton was also only possible because of your sacrifices. I love you so much, Eomma and Appa. Sooyun, I love you too.

To God: I would like lift all my praise up to God, who gave my life new purpose and filled me with His love and strength. There were highs and lows throughout my past years at Princeton, but God, you remained faithful to me. I want to live my life trusting in you and glorifying you.

*Trust in the Lord with all your heart, and do not lean on your own understanding.
In all your ways acknowledge him, and he will make straight your paths.*

Proverbs 3:5-6

Contents

Abstract	iii
Acknowledgements	iv
List of Tables	ix
List of Figures	xi
1 Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
2 Literature Review	4
2.1 Commodity Futures	4
2.2 Regime-Switching Approach	5
2.2.1 Regime Identification	6
2.2.2 Regime Prediction	8
2.3 Portfolio Construction Methods	8
2.3.1 Mean Reversion	9
2.3.2 Trend-Following	9
2.3.3 Risk Parity	10
2.4 Contributions to Existing Literature	11
3 Data	13
3.1 Description of Data	13

3.1.1	Goldman Sachs Commodity Index (GSCI)	13
3.1.2	Invesco DB Commodity Index Tracking Fund (DBC)	15
3.1.3	FRED-MD Macroeconomic Database	17
3.2	Exploratory Data Analysis	17
3.2.1	Summary Statistics	17
3.2.2	Basic Analysis	22
4	Methodology	25
4.1	Regime Identification	25
4.1.1	ℓ_1 Trend-Filtering	25
4.1.2	Jump Models	27
4.1.3	Time Series Cross-Validation	28
4.2	Regime Prediction	29
4.2.1	Logistic Regression	30
4.2.2	Random Forests	31
4.2.3	Neural Networks	32
4.3	Portfolio Construction Methods	33
4.3.1	Mean Reversion	34
4.3.2	Trend-Following	34
4.3.3	Risk-Parity	34
4.3.4	Regime Switching Framework	35
4.3.5	Backtesting Framework	36
5	Results and Analysis	41
5.1	Regime Identification Outcomes	41
5.1.1	Trend Filtering	41
5.1.2	Continuous Statistical Jump Models	46
5.2	Regime Prediction Outcomes with Trend Filtering	50

5.2.1	Logistic Regression	51
5.2.2	Neural Networks	53
5.2.3	Random Forests	53
5.2.4	Gradient Boosting	56
5.3	Regime Prediction Outcomes with Statistical Jump Models	56
5.3.1	Logistic Regression	58
5.3.2	Neural Networks	58
5.3.3	Random Forests	62
5.3.4	Gradient Boosting	62
5.4	Final Remarks on Regime Prediction	64
5.5	Portfolio Construction	65
5.5.1	Mean Reversion	65
5.5.2	Trend-Following	72
5.5.3	Risk Parity	77
6	Conclusion	84
6.1	Summary of Results	84
6.2	Limitations and Future Research	86

List of Tables

3.1	Index Base Weights of DBC (Source: Invesco)	16
3.2	Comparison of GSCI and S&P 500 Since May 1991	18
3.3	Comparison of GSCI & S&P 500 Since February 2020	19
3.4	Comparison of DBC and S&P 500 Since February 2006	20
3.5	Comparison of DBC and S&P 500 Since February 2020	22
4.1	Components of Each Portfolio	37
5.1	GSCI Jump Model Loss for Different λ Values	47
5.2	DBC Jump Model Loss for Different λ Values	48
5.3	GSCI Regime Prediction Test Accuracy with Regime Information Obtained Through Trend Filtering	50
5.4	DBC Regime Prediction Test Accuracy with Regime Information Obtained Through Trend Filtering	51
5.5	GSCI Regime Prediction Test Accuracy with Regime Information Obtained Through Jump Models	58
5.6	DBC Regime Prediction Test Accuracy with Regime Information Obtained Through Jump Models	58
5.7	Portfolio 1 Mean Reversion Performance - GSCI	66
5.8	Portfolio 1 Mean Reversion Performance - DBC	67
5.9	Portfolio 2 Mean Reversion Performance - GSCI	67

5.10 Portfolio 2 Mean Reversion Performance - DBC	68
5.11 Portfolio 3 Mean Reversion Performance - GSCI	69
5.12 Portfolio 3 Mean Reversion Performance - DBC	69
5.13 Portfolio 4 Mean Reversion Performance - GSCI	70
5.14 Portfolio 4 Mean Reversion Performance - DBC	71
5.15 Portfolio 1 Trend-Following Performance - GSCI	73
5.16 Portfolio 1 Trend-Following Performance - DBC	73
5.17 Portfolio 2 Trend-Following Performance - GSCI	74
5.18 Portfolio 2 Trend-Following Performance - DBC	74
5.19 Portfolio 3 Trend Following Performance - GSCI	75
5.20 Portfolio 3 Trend Following Performance - DBC	75
5.21 Portfolio 4 Trend Following Performance - GSCI	76
5.22 Portfolio 4 Trend Following Performance - DBC	76
5.23 Portfolio 1 Risk Parity Performance - GSCI	78
5.24 Portfolio 1 Risk Parity Performance - DBC	78
5.25 Portfolio 2 Risk Parity Performance - GSCI	79
5.26 Portfolio 2 Risk Parity Performance - DBC	80
5.27 Portfolio 3 Risk Parity Performance - GSCI	80
5.28 Portfolio 3 Risk Parity Performance - DBC	81
5.29 Portfolio 4 Risk Parity Performance - GSCI	81
5.30 Portfolio 4 Risk Parity Performance - DBC	82

List of Figures

3.1	GSCI Index Weights (Source: S&P Global)	15
3.2	GSCI Cumulative Log Returns Over Time	18
3.3	Comparison of Closing Prices for GSCI Index and S&P500 Index	18
3.4	Comparison of Cumulative Log Returns for GSCI Index and S&P500 Index	19
3.5	Cumulative Log Returns of GSCI Index and S&P 500 Index After Covid-19 Outbreak	20
3.6	DBC Cumulative Log Returns Over Time	20
3.7	Comparison of Closing Prices for DBC Index and S&P 500 Index	21
3.8	Comparison of Cumulative Log Returns for DBC Index and S&P 500 Index	21
3.9	Cumulative Log Returns of DBC Index and S&P 500 Index After Covid-19 Outbreak	22
4.1	Comparison of Cumulative Returns for Portfolio Assets	37
5.1	Comparison of Trend-filtering with Different Hyperparameter λ	42
5.2	Contraction Regimes Identified Through Trend Filtering with the Tuned λ Value	43
5.3	Contraction Regimes Identified in GSCI Data Through Trend Filtering with Yamada's Hyperparameter Tuning	44

5.4	Contraction Regimes Identified in DBC Data Through Trend Filtering with the Tuned λ Value	45
5.5	Contraction Regimes Identified in DBC Data Through Trend Filtering with Yamada's Hyperparameter Tuning	45
5.6	Contraction Regimes Identified in GSCI Data Through Jump Models	47
5.7	Probabilities of Contraction Regime in GSCI Data	48
5.8	Contraction Regimes Identified in DBC Data Through Jump Models	49
5.9	Probabilities of Contraction Regime in DBC Data	49
5.10	Contraction Regimes of GSCI Predicted Through Trend Filtering + Logistic Regression	52
5.11	Contraction Regimes of DBC Predicted Through Trend Filtering + Logistic Regression	52
5.12	Contraction Regimes of GSCI Predicted Through Trend Filtering + Feedforward Neural Networks	53
5.13	Contraction Regimes of DBC Predicted Through Trend Filtering + Feedforward Neural Networks	54
5.14	Contraction Regimes of GSCI Predicted Through Trend Filtering + Recurrent Neural Networks	54
5.15	Contraction Regimes of DBC Predicted Through Trend Filtering + Recurrent Neural Networks	55
5.16	Contraction Regimes of GSCI Predicted Through Trend Filtering + Random Forests	55
5.17	Contraction Regimes of DBC Predicted Through Trend Filtering + Random Forests	56
5.18	Contraction Regimes of GSCI Predicted Through Trend Filtering + Gradient Boosting	57

5.19	Contraction Regimes of DBC Predicted Through Trend Filtering + Gradient Boosting	57
5.20	Contraction Regimes of GSCI Predicted Through Jump Models + Logistic Regression	59
5.21	Contraction Regimes of DBC Predicted Through Jump Models + Logistic Regression	59
5.22	Contraction Regimes of GSCI Predicted Through Jump Models + Feedforward Neural Networks	60
5.23	Contraction Regimes of GSCI Predicted Through Jump Models + Recurrent Neural Networks	60
5.24	Contraction Regimes of DBC Predicted Through Jump Models + Feedforward Neural Networks	61
5.25	Contraction Regimes of DBC Predicted Through Jump Models + Recurrent Neural Networks	61
5.26	Contraction Regimes of GSCI Predicted Through Jump Models + Random Forests	62
5.27	Contraction Regimes of DBC Predicted Through Jump Models + Random Forests	63
5.28	Contraction Regimes of GSCI Predicted Through Jump Models + Gradient Boosting	63
5.29	Contraction Regimes of DBC Predicted Through Jump Models + Gradient Boosting	64

Chapter 1

Introduction

1.1 Motivation

Commodities play a pivotal role in global economics, serving not only as essential goods for consumption and inputs for production but also as instruments for investment. Despite such importance, they have received relatively little attention compared to other financial instruments such as stocks and bonds. However, due to recent market upheavals caused by events such as the outbreak of the Covid-19 pandemic and Russia's invasion of Ukraine in 2022, commodity prices have been showing extreme volatility changes, shedding light on increased need for accurate prediction and hedge against commodity price changes.

For instance, a study by Wang et al. (2022) elucidates the dramatic increase in the levels of volatility spillover in the commodity market due to the geopolitical conflict between Russia and Ukraine. Similarly, Alam et al. (2022) states that the Russian invasion has created ripple effects across not only markets for different commodities but also other financial instruments, highlighting the far-reaching economic consequences of the invasion. Also, Zhang and Wang (2022) found that the COVID-19 pandemic heightened the long-run volatility of four major commodity futures – gold,

oil, soybean and copper – with two of these commodities experiencing instantaneous volatility spikes.

The unprecedented volatility of commodity markets can be deemed as a challenge to investors, but on the other hand, can unveil investment avenues. One tool that can be used for navigating this unpredictability is commodity futures. The tactical allocation of commodity futures for portfolio diversification has proven to be beneficial due to its low correlation with other asset classes such as stocks or bonds (Jensen et al., 2002). Research shows that these instruments could generate additional return through active management such as the employment of trend-following strategies, underscoring the necessity of innovative approaches (Mulvey et al., 2004). Herein lies the potential of the regime-switching approach. This approach can offer a nuanced understanding of the market dynamics of commodity futures by identifying and forecasting economic regimes. Especially in an environment marked by the likes of geopolitical tensions and global pandemics, we aim for this approach to ultimately pave the way for optimal investment decisions with the purpose of hedging against risks and maximizing return.

1.2 Problem Statement

While research exists on the regime-switching approach for the analysis and forecasting of commodity futures prices, there exists a notable gap concerning its application in the active management of commodity futures, a gap that is more eminent in light of recent market upheavals stemming from geopolitical conflicts and the global pandemic. Moreover, there remains more to be uncovered in terms of how the forecasting of regime changes could be practically integrated into trading decisions when applied along with active management. Consequently, this thesis poses the central research question: “How can the regime-switching framework be effectively applied to port-

folio optimization including commodity futures in a way that truly capture current market dynamics and generate profit?”

In this paper, we identify distinct regimes in the commodity futures market by employing advanced regime identification techniques such as jump models or trend-filtering algorithms, and also establish a regime prediction model through machine-learning techniques such as neural networks or logistic regression. Furthermore, we use the obtained regime model to evaluate the effectiveness of different portfolios with varying components through the lens of a regime-switching approach to accurately capture the changes in market dynamics, and compare its performance against other traditional portfolios.

The goal is to ascertain the most effective method for both identifying and forecasting regime shifts in commodity futures prices. We aim to assess the compatibility between the regime-switching approach and different portfolio construction methods, using specifically the S&P Goldman Sachs Commodity Index (GSCI) and the Invesco DB Commodity Index (DBC). We implement regime-identification and prediction through a variety of advanced techniques and then assess diverse set of strategies that are optimized under the regime-switching framework.

Chapter 2

Literature Review

2.1 Commodity Futures

Commodities have been actively utilized and traded for a long time, playing a fundamental role in the global economy, not only for their material value but also for their potential as an alternative investment. Accordingly, the behavior of commodity prices have long been a topic of discussion that many investors have been attempting to understand for profitable returns. The growth in commodities trading has ultimately led to an increased interest in commodity derivatives, one of them being commodity futures.

A commodity futures contract is an agreement to buy or sell a specified quantity of commodity at a predetermined price on a set date (Gorton & Rouwenhorst, 2006). Commodity futures have been considered as excellent portfolio diversifiers due to its low correlation with other financial assets, and they also serve as hedging tools against inflation, given their close connection to the government's monetary policy (Cheung & Miu, 2010; Erb & Harvey, 2006). Commodity futures have traditionally been subject to passive management, predominantly the buy-and-hold strategy. Yet, research suggests that the active management of commodity futures through effective

strategies could yield superior performance. For instance, Erb and Harvey (2006) analyzed the historical performance of commodity futures and proved that leveraging momentum-based strategies or taking advantage of information on term structures have generated attractive returns. Similarly, Miffre and Rallis (2007) examined 56 momentum and contrarian strategies and identify 13 momentum strategies that led to high profitability. These findings underscore the importance of active management of commodity futures to optimize the investment and generate profits comparable to those of other financial assets such as equities.

2.2 Regime-Switching Approach

However, with market disruptions caused by events such as the Covid-19 pandemic and the Russian invasion of Ukraine, even portfolio optimization strategies may not be sufficient to accurately capture the price patterns of commodity futures and exploit such pattern. Studies have shown that since the Covid-19 pandemic, the relationships between different commodities as well as the price dynamics of the commodity futures have changed. Borgards et al. (2021) conducted an analysis of 20 commodity futures prices after the pandemic to measure price overreactions, a large log price change between two turning points followed by a price reversal. They concluded that although the degree of overreaction differs by the type of commodity, the magnitude and frequency of price overreactions have increased during the first wave of the pandemic, further suggesting that investors should be able to frequently monitor the price dynamics in order to exploit such overreactions and prevent huge losses. Zhang and Wang (2022) also demonstrated that the four major commodity futures all experienced change in long-run volatility after the strike of the pandemic. The Russian invasion of Ukraine is another global event that spurred extreme volatility changes in commodity prices, leading to unpredictability of futures prices. For instance, Wang

et al. (2022) showed that volatility spillover spiked from 35% to 85% after the Russian invasion, a level of impact even higher than the pandemic, and Alam et al. (2022) stated that the Russian invasion has increased connectedness across not only markets for different commodities but also other financial instruments, highlighting the market contagion during this war crisis as well as the underlying systematic risk.

Whenever financial assets experience a drastic change, they display a pattern of contagion in which the changes such as increase in volatility or sharp losses persist for a period of time. Such periods in which the financial market display similar performance and common characteristics are called regimes. The aforementioned findings suggest that investors must be able to understand and even predict the changes in the dynamics of commodity prices and highlight the importance of implementing a regime-switching framework that integrates shifts in market patterns caused by geopolitical events, policy changes, etc.

Choi and Hammoudeh (2010) revealed that low-volatility and high-volatility regimes are observed in the five strategic commodity prices and US S&P 500 index, and that in each regime, the optimal set of actions for policymakers and investors may vary. Beckmann et al. (2014) examined the spot and futures prices of energy commodities and concluded that a non-linear adjustment in the spot and futures prices under a smooth regime-switching framework is needed to accurately capture the complex dynamics of commodity spots and futures.

2.2.1 Regime Identification

In light of such growing importance in identifying and predicting regime shifts, researchers have been concentrating on more adaptive and sophisticated models. The concept of identifying regimes from financial time series was first introduced by Hamilton (2010), when he applied the Markov switching model to the U.S. real Gross National Product (GNP) to define growth cycles and recession regimes that turned

out to be consistent with the business cycles identified by the National Bureau of Economic Research (NBER).

Since Hamilton's initial regime-switching approach, various efforts have been made to precisely identify regimes. One such method involves Hamilton's approach to use hidden Markov models. For instance, Liew and Siu (2010) applied the hidden Markov model for regime-dependent option valuation, and M.-J. Kim and Yoo (1995) broadened the application of the univariate hidden Markov model to a multivariate model. Additionally, Assoe (1998) utilized the hidden Markov model to demonstrate the regime-changing behavior in stock market returns. Bae et al. (2014) constructed a hidden Markov model framework to identify the regimes from time series data of stock, bond and commodity market indices, leveraging the framework for portfolio optimization.

A generalized form of hidden Markov models (HMMs) would be jump models, proposed by Nystrup et al. (2020). Traditional HMMs often encounter difficulties in accurately capturing market regimes due to their short-lived state persistence. Jump models introduce penalties for state transitions in order to increase the persistence of states, thereby better reflecting the true underlying regimes and providing a more robust estimation of market dynamics. Aydinhan et al. (2023) further advanced this concept through the development of continuous statistical jump models. Transitioning from the identification of discrete hidden regime states to the utilization of a probability vector, this approach proves particularly beneficial in active risk management, as it enhances the ability to detect market downturns, even in periods of apparent market stability.

Another approach for regime identification is the application of trend filtering. Mulvey and Liu (2016) demonstrated that trend-filtering algorithms can extract piecewise linear trend components to identify economic regimes effectively.

In our study, we employ both the trend-filtering algorithm and continuous sta-

tistical jump models to assess and compare their performance in identifying regimes within commodity futures data. Through this comparative analysis, we determine the method that can appropriately capture the regimes observed from the commodity futures data, providing a solid foundation for the subsequent task of regime prediction.

2.2.2 Regime Prediction

The regime identification task is followed by regime prediction, a task that yields significant advantages for investors and policymakers alike by facilitating informed decision-making in anticipation of changes in market dynamics. Recent endeavors have explored the usage of advanced machine learning techniques to predict regimes. For instance, Uysal and Mulvey (2021) employed supervised models to learn the relationship between market features and the regime labels, and their findings demonstrated how a regime-based risk parity framework could enhance portfolio performance.

Despite these recent advancements in research, the application of machine learning for regime prediction for active management remains a relatively unexplored area, given the scarcity of research in the methodology or empirical outcomes of applying machine learning models for predictive purposes in this context. This gap presents a unique opportunity to explore the efficacy of machine learning techniques in predicting regimes within the unpredictable realm of commodity futures trading.

2.3 Portfolio Construction Methods

A diverse set of portfolio construction methods have been proposed with the purpose of enhanced performance and risk management. Some notable methods are mean reversion, trend-following, risk parity. These methods have been widely adopted by investors based on market conditions or specific investment goals. This section reviews

existing literature on the aforementioned methods, highlighting their application, advantages, and limitations, along with a focus on their relevance to the regime-switching approach.

2.3.1 Mean Reversion

The mean reversion strategy leverages the assumption that asset prices and historical returns eventually revert to their long-term average level to predict future price movements. The empirical evidence of mean reversion in the U.S. stock markets was first suggested by Poterba and Summers (1988), and Fama and French (1988). Since then, this strategy has garnered attention as an effective strategy that enables long-term investors to identify and invest in assets that have deviated from their historical average.

Mean reversion can be utilized within a regime-switching framework, adding sophistication to the strategy by enabling investors to adjust the strategy based on the prevailing market regime. For instance, Nielsen and Olesen (2001) fitted the two-state regime-switching model to the Danish stock returns, observing different levels of mean reversion in the respective regimes. Several studies have posited that there is not statistically significant evidence of mean reversion present in commodity futures under a single-regime approach, but such varying degrees of mean reversion in different regimes suggest that mean reversion strategy might be applicable to commodity futures under a regime-switching approach (Andersson, 2007; Irwin et al., 1996).

2.3.2 Trend-Following

Trend-following strategies capitalize on persistent patterns observed in asset prices. They have proven to show strong performance across diverse asset classes and economic environments, providing diversification benefits and downside protection, especially during market downturns (Hurst et al., 2017; Moskowitz et al., 2012). These

strategies are particularly relevant in the context of commodity futures, where price movements can be significantly influenced by factors such as seasonal patterns, geopolitical events, and changes in supply and demand dynamics. Mulvey et al. (2004) demonstrated that the MLM index, a trend-following commodity futures index, is a useful tool for portfolio diversification. Likewise, Szakmary et al. (2010) examined the profitability of the implementation of trend-following strategies to long-term futures for 28 commodity markets over a 48-year period, discovering that pure trend-following strategies yield profits that are higher than both passive management and momentum strategies. Overall, these findings show that trend-following strategies can enhance the performance of commodity futures, and many commodity trading advisors (CTAs) have in fact been actively implementing these strategies for trading these alternative assets.

Although trend-following strategies have proven to be profitable, some criticize the method's reliance on historical price movements to predict future trends, which oversimplifies the dynamic nature of the financial markets and may lead to false signals when there are rapid market changes. However, we believe that the integration of a regime-switching framework can effectively address these concerns, enhancing the robustness and adaptability of trend-following strategies in varying market regimes.

2.3.3 Risk Parity

The risk parity approach to portfolio construction emphasizes equalizing risk contributions from all assets within a portfolio, rather than allocating capital based on market capitalization or expected returns, ultimately placing risk diversification above capital diversification (Qian, 2011). This method has advantages of higher risk-adjusted returns and is also suitable for managing portfolios of assets with diverse risk levels. One potential disadvantage is that the portfolio will leverage low-risk assets to counteract high-risk assets, creating an imbalance in the portfolio.

A risk parity portfolio can be even more beneficial under the regime-switching approach; Costa and Kwon (2019) explored the risk parity portfolio under a Markov regime-switching framework and concluded that a two-state regime-switching portfolio can consistently provide higher-than-nominal risk-adjusted returns over long investment horizons, due to better quality estimates of the asset covariance matrix. The regime-switching risk parity portfolio can be especially beneficial for portfolios including commodity futures, considering the inherently volatile nature of commodities markets.

2.4 Contributions to Existing Literature

While machine learning has been increasingly applied in financial markets, there is a notable gap in applying these techniques to regime identification and prediction in financial assets, particularly in the realms of commodity futures. Notably, prior research has concentrated on applying the regime-switching framework to traditional financial assets like equities or relegating machine learning to broader contexts within finance. The amalgamation of a regime-switching framework with machine learning methodologies to capture commodity futures market dynamics thus remains relatively unexplored. Furthermore, although previous studies have shed light on the potential of active management of commodity futures to exploit systematic profit opportunities in futures markets, there are few that focus on how different portfolio construction methods can be optimized with the regime-switching approach, especially after global events such as the the Covid-19 pandemic and the Russian invasion of Ukraine that could significantly impact the market dynamics.

In response, we seek to integrate the regime-switching framework with machine learning algorithms for a nuanced understanding of the intricate commodity futures dynamics. We employ a blend of trend-filtering algorithm, continuous statistical jump

models and machine-learning methods such as neural networks, random forests, gradient boosting and logistic regression models for regime identification and prediction. Furthermore, we use the acquired regime information to propose a novel method of regime-switching portfolio optimization. By doing so, we aim to contribute a novel perspective to the literature, offering practical solutions and strategic insights for investors and policymakers in the complex world of commodity futures trading.

Chapter 3

Data

3.1 Description of Data

In this section, we describe the sources and characteristics of the data we use in this research. Furthermore, we conduct an exploratory data analysis to acquire a preliminary understanding of the dataset before its application in the main analysis.

3.1.1 Goldman Sachs Commodity Index (GSCI)

Our main dataset is the market data from the S&P Goldman Sachs Commodity Index (GSCI). GSCI, the first major investable commodity index launched in May 1991, is one of the most widely traded commodity indices that provides diversified exposure to global commodities. The index consists of 24 exchange-traded futures from the commodity sectors of energy products, industrial metals, agricultural products, live-stock products and precious metals. GSCI reflects a passive portfolio of futures with long-only positions, and thus could be considered as a portfolio of futures managed with passive buy-and-hold strategy.

GSCI is weighted based on world-production levels, ensuring that each commodity's representation within the index is proportionate to the amount of production for

commodities. This index encompasses a wide range of commodities, encapsulating the diverse sectors within the commodities market. Consequently, the GSCI serves as a good global economic indicator as well as a measure for investment performance, capturing the supply and demand dynamics within the commodity markets.

The GSCI index is rebalanced annually in January with respect to contracts that have trading volume multiple (TVM) below the TVM Threshold. The contract production weight (CPW) of all eligible futures contracts are determined, and the compositions of the index are adjusted accordingly.

One reason for choosing GSCI is because of its availability of data. As the first investable commodity index, the GSCI has a long history, with data available since its inception in 1991. This extensive data set is valuable to understand the historical behaviors of the commodities market and to identify and predict regimes from the observed market behaviors. Moreover, the historical data of GSCI is available from major financial data providers, making it a reliable source for research and analysis. Another reason is due to its liquidity as one of the most widely traded indices, being less susceptible to large, erratic price swings caused by low volume trading that could skew the analysis outcomes. Lastly, since the fund consists of futures contracts, we believed it was relevant to the objective of our research which is contingent to commodity futures.

We extract the data for GSCI from Yahoo Finance using the Python *yfinance* library, spanning from May 1991 to March 2024. We choose Yahoo Finance as the database due to its compatibility with Python, the programming language that we mainly use for analysis. The dataset consists of daily open, high, low and closing prices, and trading volumes. We remove any columns with missing values to ensure data integrity and consistency throughout our analysis.

Table 1: Contracts Included in the S&P GSCI

Trading Facility	Commodity	Ticker ⁽¹⁾	2022 CPW	2023 CPW	2022 ACRP (\$)	Unit	2022 RPDW ⁽²⁾	Designated Contract												
								2023 RPDW	1	2	3	4	5	6	7	8	9	10	11	12
CBT	Chicago Wheat	W	19458.43	19263.74	8.686666667	bu	3.64%	3.31%	H	H	K	K	N	N	U	U	Z	Z	Z	H
KBT	Kansas Wheat	KW	7915.366	8365.251	9.096041667	bu	1.40%	1.50%	H	H	K	K	N	N	U	U	Z	Z	Z	H
CBT	Corn	C	43170.92	43970.54	6.511458333	bu	6.54%	5.66%	H	H	K	K	N	N	U	U	Z	Z	Z	H
CBT	Soybeans	S	12270.07	12491.57	14.60375	bu	4.64%	3.60%	H	H	K	K	N	N	X	X	X	X	F	F
ICE - US	Coffee	KC	20672.35	21127.24	2.238875	lbs	0.83%	0.93%	H	H	K	K	N	N	U	U	Z	Z	Z	H
ICE - US	Sugar	SB	390154.8	391575.0	0.187483333	lbs	1.81%	1.45%	H	H	K	K	N	N	V	V	V	H	H	H
ICE - US	Cocoa	CC	5.013083	5.170340	2492.333333	MT	0.36%	0.25%	H	H	K	K	N	N	U	U	Z	Z	Z	H
ICE - US	Cotton	CT	54201.50	54332.83	1.179533333	lbs	1.26%	1.27%	H	H	K	K	N	N	Z	Z	Z	Z	Z	H
CME	Lean Hogs	LH	94540.48	96620.79	0.956645833	lbs	2.36%	1.83%	G	J	J	M	M	N	Q	V	V	Z	Z	G
CME	Live Cattle	LC	110644.1	110633.0	1.363395833	lbs	3.76%	2.98%	G	J	J	M	M	Q	Q	V	V	Z	Z	G
CME	Feeder Cattle	FC	29541.00	31753.94	1.672854167	lbs	1.25%	1.05%	H	H	J	K	Q	Q	Q	U	V	X	F	F
NYM / ICE	WTI Crude Oil	CL	12392.85	12079.15	91.4525	bbl	20.34%	21.83%	G	H	J	K	M	N	Q	U	V	X	Z	F
NYM	Heating Oil	HO	70902.91	75458.02	3.100816667	gal	3.50%	4.62%	G	H	J	K	M	N	Q	U	V	X	Z	F
NYM	RBOB Gasoline	RB	85545.41	82553.26	2.817091667	gal	4.34%	4.60%	G	H	J	K	M	N	Q	U	V	X	Z	F
ICE - UK	Brent Crude Oil	LCO	10024.82	10693.69	94.36333333	bbl	17.19%	19.94%	H	J	K	M	N	Q	U	V	X	Z	F	G
ICE - UK	Gasoil	LGO	342.1020	319.4099	912.6041667	MT	4.78%	5.76%	G	H	J	K	M	N	Q	U	V	X	Z	F
NYM / ICE	Natural Gas	NG	37653.94	39421.14	6.056416667	MMBtu	3.33%	4.72%	G	H	J	K	M	N	Q	U	V	X	Z	F
LME	Aluminum	MAL	65.89200	67.89000	2835.175	MT	4.18%	3.80%	G	H	J	K	M	N	Q	U	V	X	Z	F
LME	Copper	MCU	23.58000	23.84000	9230.270833	MT	5.80%	4.35%	G	H	J	K	M	N	Q	U	V	X	Z	F
LME	Nickel	MNI	2.016000	2.093800	23762.83333	MT	1.00%	0.98%	G	H	J	K	M	N	Q	U	V	X	Z	F
LME	Lead	MPB	11.02000	11.22000	2211.320833	MT	0.66%	0.49%	G	H	J	K	M	N	Q	U	V	X	Z	F
LME	Zinc	MZN	13.44000	13.52000	3560.584167	MT	1.08%	0.95%	G	H	J	K	M	N	Q	U	V	X	Z	F
CMX	Gold	GC	102.3680	103.7183	1822.708333	oz	5.33%	3.74%	G	J	J	M	M	Q	Q	Z	Z	Z	Z	G
CMX	Silver	SI	887.3607	879.0015	22.27158333	oz	0.64%	0.39%	H	H	K	K	N	N	U	U	Z	Z	Z	H

Figure 3.1: GSCI Index Weights (Source: S&P Global)

3.1.2 Invesco DB Commodity Index Tracking Fund (DBC)

Another dataset we use is the market data from Invesco DB Commodity Index Tracking Fund (DBC). DBC is a widely traded exchange-traded fund (ETF) that provides exposure to futures contracts of 14 of the most heavily traded commodities of the world (Light Sweet Crude Oil (WTI), Heating Oil, RBOB Gasoline, Natural Gas, Brent Crude, Gold, Silver, Aluminum, Zinc, Copper Grade A, Corn, Wheat, Soybeans, and Sugar). It tracks changes in the level of the DBIQ Optimum Yield Diversified Commodity Index Excess Return (DBIQ Opt Yield Diversified Commodity

Index ER) and encapsulates the overall commodity market dynamics with diversified exposures to major commodities.

One reason for choosing DBC is because it offers a diversified basket of commodities, ensuring that the analysis is not skewed by the performance of a single commodity and ultimately offering a more comprehensive view of the commodity markets. Another reason is due to its liquidity as one of the most widely traded ETFs, similar to the rationale of choosing GSCI. Also, the availability of data was again a significant factor; we were able to obtain DBC's historical data through Yahoo Finance.

The DBC index is rebalanced annually in November to maintain the base weights in the same proportion of the base weights of commodities on the base date (September 3, 1997). The base weights are shown in Table 3.1.2.

Index Commodity	Base Weight (%)
Light Sweet Crude Oil (WTI)	12.375
Heating Oil	12.375
RBOB Gasoline	12.375
Natural Gas	5.500
Brent Crude	12.375
Gold	8.000
Silver	2.000
Aluminium	4.167
Zinc	4.167
Copper Grade A	4.167
Corn	5.625
Wheat	5.625
Soybeans	5.625
Sugar	5.625

Table 3.1: Index Base Weights of DBC (Source: Invesco)

The data for DBC spans from February 2006, its inception date, to March 2024. The dataset consists of daily open, high, low and closing prices, and trading volumes. As with the GSCI, we remove any columns with missing values to ensure the quality of data for our analysis.

3.1.3 FRED-MD Macroeconomic Database

Lastly, we use the FRED-MD macroeconomic database compiled by McCracken and Ng (2016). Specifically, we use the 134 macroeconomic feature variables in the database in order to perform regime prediction. We incorporate the 134 features as predictor variables when predicting the regime for the commodity futures returns series using machine learning methods such as logistic regression or random forests. The database includes variables such as the effective federal funds rate, unemployment rates, real estate loans, etc., and is updated monthly.

3.2 Exploratory Data Analysis

3.2.1 Summary Statistics

After initial extraction, we compute the cumulative log returns from the adjusted close prices, as log returns offer a normalized measure of performance for analysis across different time frames. See Figure 3.2 for the plot for GSCI's cumulative log return values. Additionally, the annualized return and the annual standard deviation are computed as summary statistics for an enhanced understanding of the index's long-term performance.

Furthermore, we compare the GSCI index with the S&P 500, in order to understand the different natures or interdependence of commodity markets and equity markets, with the consideration that the S&P 500 index is an equity benchmark that reflects the equity performance of the biggest companies in the United States. See Figure 3.3 and Table 3.2 for the specific comparison outcomes since the GSCI's inception in May 1991. The correlation between GSCI and S&P 500 is computed as the value of 0.2218.

Then, we compute the statistics of both S&P 500 and GSCI for the time period

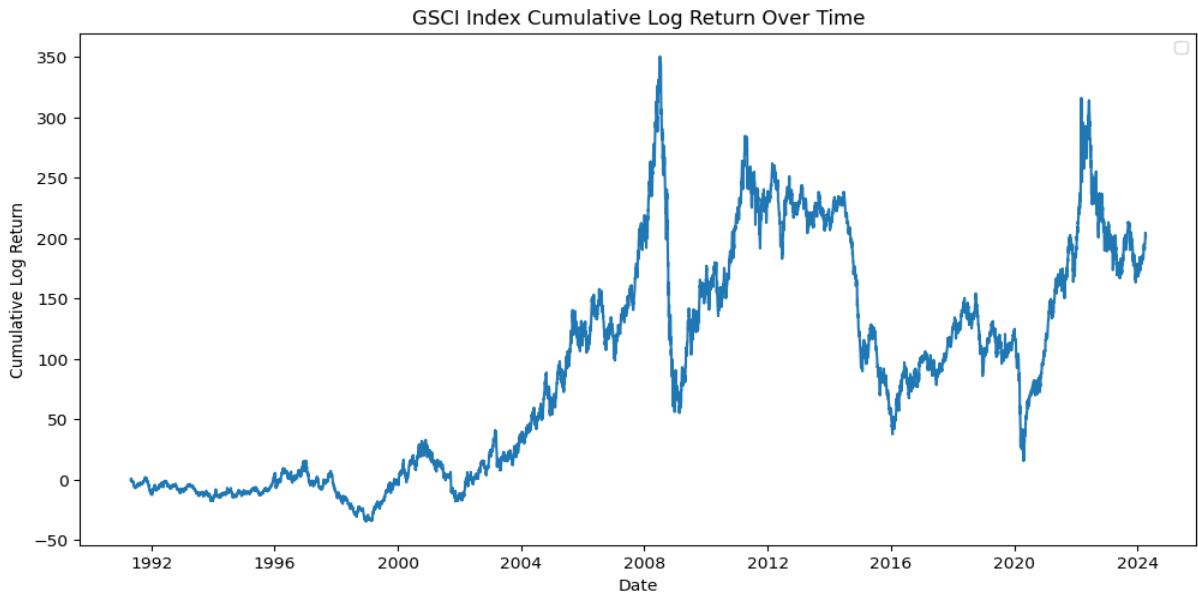


Figure 3.2: GSCI Cumulative Log Returns Over Time



Figure 3.3: Comparison of Closing Prices for GSCI Index and S&P500 Index

Index	Annualized Return	Annual Standard Deviation
GSCI	3.03 %	21.50%
S&P500	8.02 %	18.26%

Table 3.2: Comparison of GSCI and S&P 500 Since May 1991

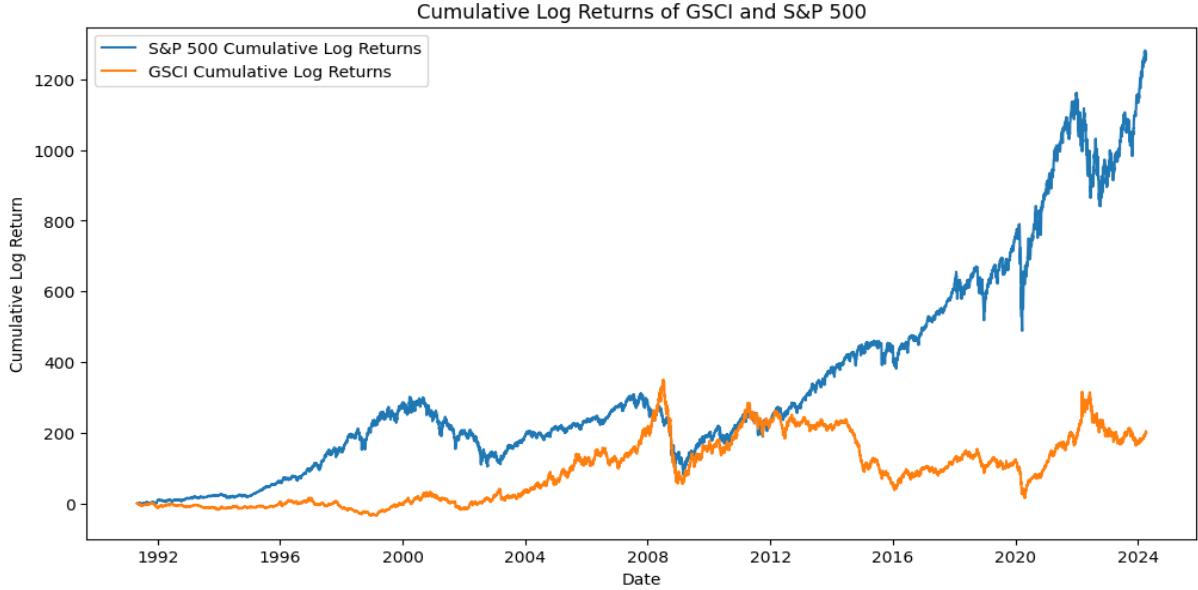


Figure 3.4: Comparison of Cumulative Log Returns for GSCI Index and S&P500 Index

after February 2020, the outbreak of the Covid-19 pandemic, to understand how the market disruptions caused by the pandemic shifted the dynamics of both assets as well as the correlation between the two assets. The annualized return and standard deviation are shown in Table 3.3. The correlation of the returns of the two assets after the outbreak of the pandemic is computed as 0.3186.

Index	Annualized Return	Annual Standard Deviation
GSCI	8.52 %	26.92%
S&P500	10.09 %	23.31%

Table 3.3: Comparison of GSCI & S&P 500 Since February 2020

Likewise, we conduct analysis on DBC data. See Figure 3.6 for a plot of the DBC index's cumulative log return values. We compute statistics to compare the DBC index with the S&P 500. See Figure 3.7 and Table 3.2.1 for the specific comparison outcomes. The annual correlation between DBC and S&P 500 is computed as the value of 0.4122.

Again, we compute the summary statistics of the two assets for the post-pandemic

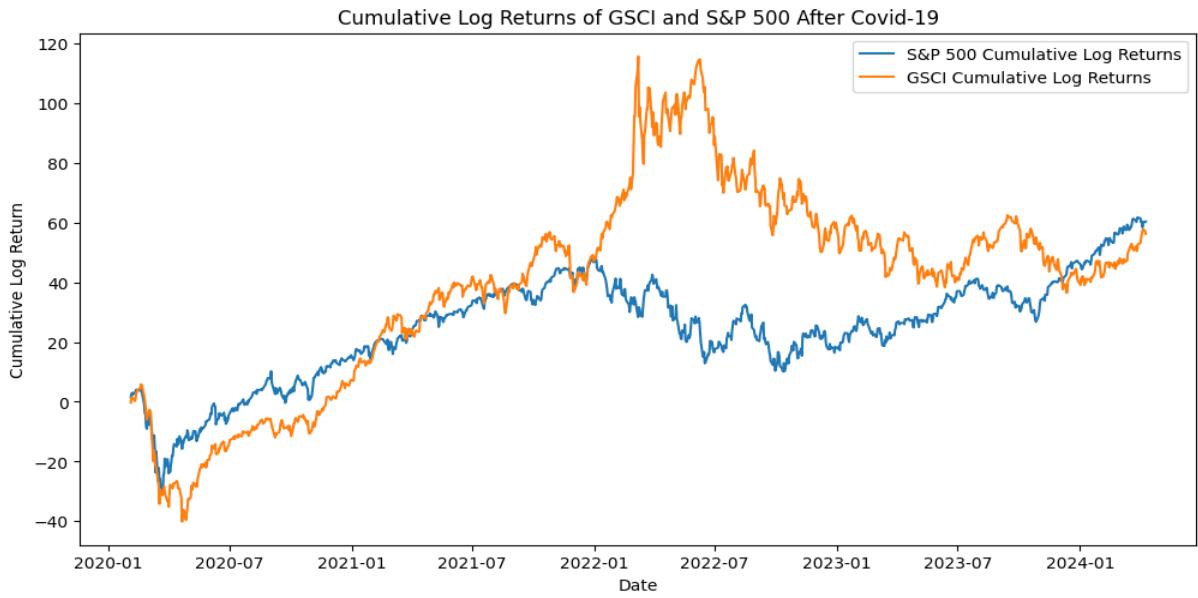


Figure 3.5: Cumulative Log Returns of GSCI Index and S&P 500 Index After Covid-19 Outbreak

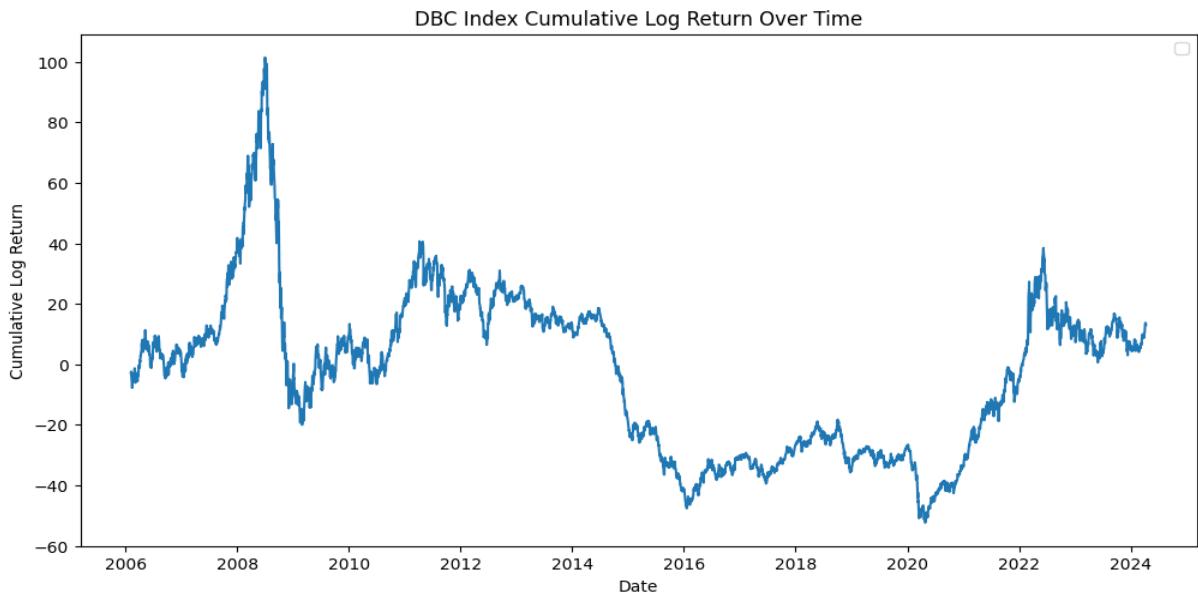


Figure 3.6: DBC Cumulative Log Returns Over Time

Index	Annual Mean Return	Annual Standard Deviation
DBC	0.23 %	19.63%
S&P500	7.64%	19.89%

Table 3.4: Comparison of DBC and S&P 500 Since February 2006



Figure 3.7: Comparison of Closing Prices for DBC Index and S&P 500 Index

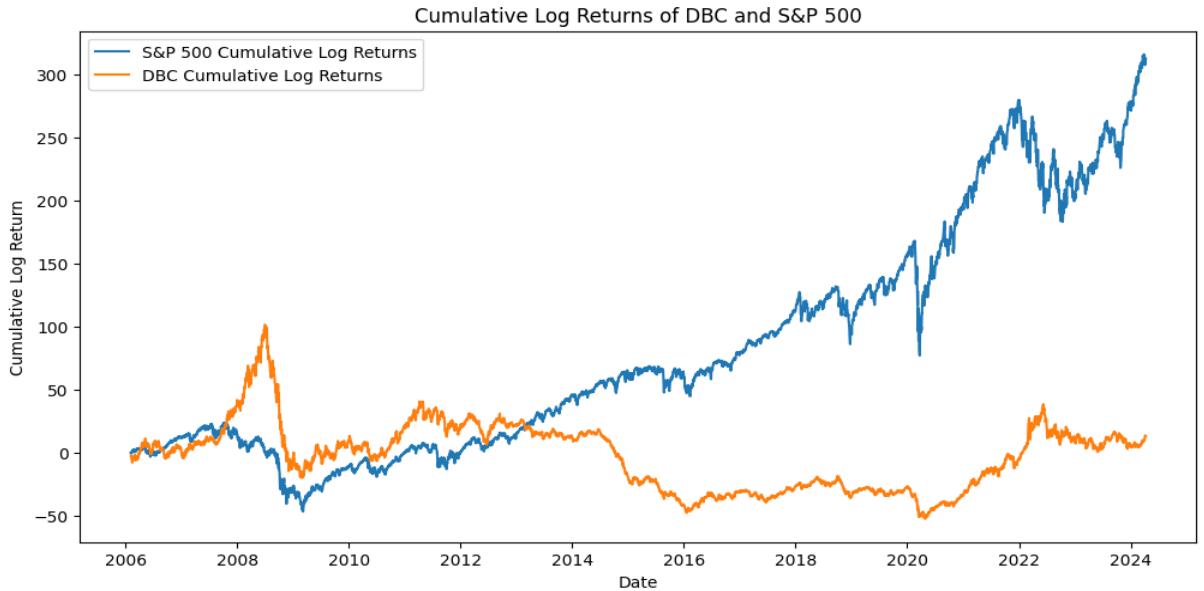


Figure 3.8: Comparison of Cumulative Log Returns for DBC Index and S&P 500 Index

time period. See Table 3.2.1 and Figure 3.9 for more details.

The correlation between the returns of DBC and S&P 500 after the pandemic is computed as 0.3669.

Index	Annual Mean Return	Annual Standard Deviation
DBC	13.05 %	21.43%
S&P500	10.09 %	23.31 %

Table 3.5: Comparison of DBC and S&P 500 Since February 2020

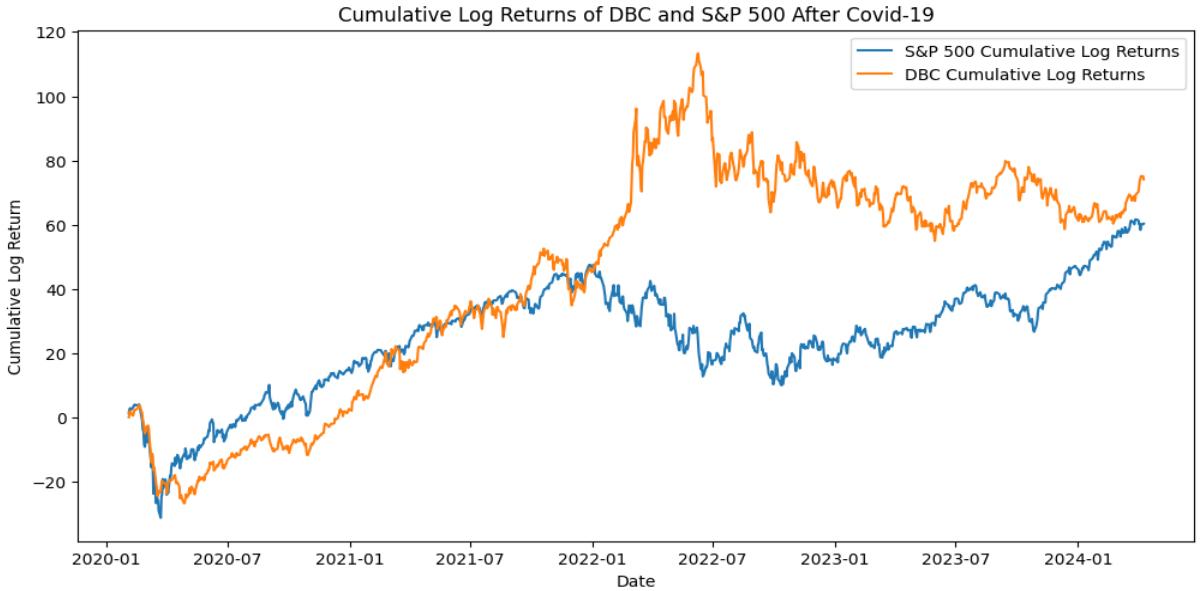


Figure 3.9: Cumulative Log Returns of DBC Index and S&P 500 Index After Covid-19 Outbreak

3.2.2 Basic Analysis

From the summary statistics that we compute, we extract some insights on how these different assets perform and interact with one another. Also, we aim to understand the changes in these different indices after the outbreak of the Covid-19 pandemic (February 2020).

Both GSCI and S&P 500 indices see significant increases in their annualized returns post-pandemic compared to the long-term trend since the inception of GSCI. This implies that there was a steady growth phase after the initial shock of the pandemic. The annual standard deviation also increases for both indices post-pandemic, indicating higher volatility after the Covid-19 outbreak. The S&P 500 shows higher annualized returns than the GSCI both the long-term (since May 1991) and post-

pandemic periods, yet the GSCI has higher volatility in the post-pandemic period. This could be attributed to the commodity market's sensitivity to global economic disruptions and supply chain issues during the pandemic due to the market nature of dealing with physical commodities.

The correlation of 0.3186 between GSCI and post-pandemic S&P 500 returns suggests that although there exists some common market drivers and dependency, the two assets are not perfectly correlated. This further opens up the possibility that the GSCI could be used as a diversifier for an optimal portfolio.

The DBC index also shows an increase in both return and volatility post-pandemic. One notable observation is that while DBC's return was lower before the outbreak of the pandemic, its return exceeds that of the S&P 500 afterwards, suggesting its heightened potential as a profitable investment tool. The correlation between DBC and S&P 500 is slightly higher, both the correlation since the inception of DBC and the correlation after the pandemic, than that between GSCI and S&P 500, implying that DBC exhibits a more synchronized response to the market conditions with S&P 500.

In summary, the statistics reveal a notable shift in the performance dynamics of both GSCI and DBC after the pandemic; there were notable increases in both returns and volatility. Also, while the GSCI shows strong performance before the pandemic, DBC emerges as the stronger performer in the higher-volatility market following the outbreak of COVID-19.

Such distinct shifts in the performance and risk profiles of the GSCI and DBC indices before and after the outbreak of the COVID-19 pandemic suggest that traditional single-regime models may not adequately capture the complexities and dynamics inherent in commodity markets, particularly during significant economic disruptions. Our analysis supports the necessity of adopting a regime-switching approach. Such approach enables the identification and adjustment to different market states

or ‘regimes’, thus more accurately reflecting the evolving risk-return trade-offs and correlations during major economic events. We believe that this approach will therefore offer a more effective tool for investors, enhancing decision-making processes in uncertain market conditions.

Chapter 4

Methodology

4.1 Regime Identification

4.1.1 ℓ_1 Trend-Filtering

The ℓ_1 trend-filtering algorithm is a nonparametric machine-learning method widely used to estimate underlying trends in time series data, first proposed by S.-J. Kim et al. (2009). The method deconstructs a time series component y_t into a trend estimate x_t and a stochastic component z_t and aims to estimate x_t (equivalently, $z_t = y_t - x_t$).

In order to extract a smooth trend component x_t , the method makes use of the ℓ_1 norm, the sum of absolute values of the vector components. Given time series data $y \in \mathbb{R}^n$, and the trend component $x \in \mathbb{R}^n$, the trend-filtering algorithm produces an objective as follows:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{t=1}^n (y_t - x_t)^2 + \lambda \sum_{t=2}^n |x_t - x_{t-1}|.$$

The first summed term in the objective aims to minimize the residual $y_t - x_t$ between the trend component and the actual observed data, and the second summed term ensures that the trend component x_t is smoothed. λ is a hyperparameter used

to control the tradeoff between the residual and the smoothness of x_t . The algorithm produces piecewise linear trend estimates, thus well suiting time series data that has underlying piecewise linear trends.

The ℓ_1 trend-filtering algorithm has been used in various disciplines, but is most adapted to detect trends from financial data. Since financial data tends to be noisy, a smoothed trend filtering is often required in order to obtain a trend component, which further can be used to detect regimes. The strength of ℓ_1 trend-filtering is that the algorithm does not depend on any assumptions of the underlying model or stochastic processes, which could potentially lead to a divergence of outcomes depending on which model to adopt, but rather purely uses the data itself to exploit the trends.

The usage of ℓ_1 trend-filtering to identify regimes from financial time series data was first proposed by Mulvey and Liu (2016) in their paper “Identifying Economic Regimes: Reducing Downside Risks for University Endowments and Foundations.” They applied ℓ_1 trend-filtering to divide the market return time series data into periods of increases and declines, and accordingly labeled these periods as “growth” regimes and “contraction” regimes.

Similarly, we apply ℓ_1 trend-filtering in order to label the economic regimes in the commodity index time series data in this research. We define two regimes, the “growth” regime and the “contraction” regime. After we extract the regime labels from the data, we use the outcomes to further conduct regime prediction tasks.

In order to implement ℓ_1 trend filtering, we create a Python function that extracts the trend component from the input data and returns the smoothed prices in a way that minimizes the defined objective for ℓ_1 trend filtering. We use Python’s *cvxpy* library for optimization.

4.1.2 Jump Models

Jump models are a general class of latent variable models useful for describing systems characterized by abrupt, persistent changes of state dynamics (Nystrup et al., 2020). Jump models can be considered as an extension of the hidden Markov model (HMM)’s classical, order-independent clustering approaches. By incorporating temporal information, jump models can better model the regime information present in time series data compared to other clustering approaches such as K-means clustering. Also the jump model has more descriptive capabilities than HMMs, which may be considered as a very specific case of jump models (Bemporad et al., 2018).

In this research, we implement the continuous statistical jump model proposed by Aydinhan et al. (2023) as an extension of Bemporad et al. (2018)’s statistical jump model, in order to fit the state sequences to the time series data while also obtaining the probabilistic interpretation of the regime identification outcomes. Such probabilistic interpretation provides more insights than a discrete approach that simply assigns one label for a given time period.

In this case, given the observation sequence of D (standardized) features $\mathbf{Y} := \{\mathbf{y}_0, \dots, \mathbf{y}_{T-1}\}$ for $\mathbf{y}_t \in \mathbb{R}^D$, we estimate a statistical jump model with K states by solving the optimization problem

$$\arg \min_{\Theta, \mathbf{s}} \sum_{t=0}^{T-1} L(\mathbf{y}_t, \Theta, \mathbf{s}_t) + \frac{\lambda}{4} \sum_{t=0}^{T-1} \|\mathbf{s}_{t-1} - \mathbf{s}_t\|_1^2, \quad (4.1)$$

where $\mathbf{S} := [\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{T-1}]^T \in \mathbb{R}^{T \times K}$ is the hidden state vector sequence that models the probabilities across all K states, satisfying the constraints $\mathbf{S} \geq 0$ and $\mathbf{S} \mathbf{1}_K = \mathbf{1}_T$, $\Theta \in \mathbb{R}^D$ represents the model parameters, and λ is the hyperparameter for penalizing “jumps” in state sequences. Also, the loss function $L(\mathbf{y}_t, \Theta, \mathbf{s}_t)$ is defined as

$$L(\mathbf{y}_t, \Theta, \mathbf{s}_t) = \sum_k s_{t,k} l(\mathbf{y}_t, \theta_k), \quad (4.2)$$

where $l(\mathbf{y}_t, \theta_k)$ is the scaled squared ℓ_2 distance norm.

We adopt the coordinate descent algorithm approach to alternate between fitting the state sequence and the model parameters with a Quadratic Programming (QP) solver using the *gurobipy* library, Gurobi Optimization’s Python-specific library for optimization. As with ℓ_1 trend-filtering, we utilize the model for regime prediction with the goal of obtaining regime labels before conducting regime prediction through a diverse set of machine learning algorithms.

4.1.3 Time Series Cross-Validation

Time series cross-validation is a statistical technique designed to evaluate machine-learning models by dividing the time series data into the training and validation set. Unlike traditional cross-validation methods that shuffle the data before dividing into k-folds, time series cross-validation maintains the order of the data by conducting validation in a rolling basis.

The reason for using time series cross-validation is to prevent look-ahead bias, a bias that occurs when training a model using future data that wouldn’t have been available in reality. This problem commonly occurs by splitting the dataset into a training and validation set, if future data points that are unknown at time of prediction are randomly mixed and incorporated in training. This blurs the purpose of forecasting regimes in the time series data, which is done under the assumption that future information is inaccessible. By preserving the temporal structure of the data, we aim to create a robust model that can be generalized to real world observations.

We implement time series cross-validation by partitioning the data into training sets and validation sets with an expanding window, where each set contains data points in a chronological order. Then, for each fold, we train the jump model on the training set for model-fitting and hyperparameter tuning, and then apply the model to the subsequent validation set to forecast regimes without access to the future data. We

subsequently compare the performance of the model against the actual observations in the validation set using performance metrics that depend on the model. For instance, for regime identification we compare the values of the objective functions for the respective optimization problems to evaluate the model performance, and for regime prediction we look into the accuracy rate.

We additionally use time series cross validation as a means for hyperparameter tuning, specifically for the regime identification task using jump models. The performance of the models is variable depending on the hyperparameter λ that represents the degree of penalty for regime shifts. We iterate over a range of parameter values and evaluate the model performance over different folds so that we could find the optimal hyperparameter that appropriately forecasts the regime information.

In conclusion, we attempt to mitigate look-ahead bias and assess the performance of the regime identification and prediction models as we adhere to the temporal structure of the financial time series data through time series cross-validation. By doing so, we ultimately enhance the model's effectiveness in capturing regime shifts under varying market conditions.

4.2 Regime Prediction

The following subsections outline the different machine learning techniques that we use to predict the regimes from the returns data of GSCI and DBC. The general framework is to use the clustering or classification approach of these different models in order to predict the regime label of the time series data, although the details of the theoretical background and implementation of models may vary. In order to prevent look-ahead bias, we use time series cross-validation, as outlined in Section 4.1.3, to train the respective models and make predictions while preserving the temporal order of the time series. We use Python's *Scikit-learn* library to implement logistic

regression, gradient boosting and random forests, and the *Keras* library for neural networks.

4.2.1 Logistic Regression

Logistic regression is a supervised machine-learning model often used for classification tasks. Since regime prediction can be considered as classifying the data into regime labels, we determine that logistic regression would be suitable for this specific task. As we are using a two-regime approach, we focus on logistic regression models for binary classification. The model outputs the respective probability of an input $\mathbf{x} \in \mathbb{R}^m$ being classified as the $y = +1$ (growth regime) or $y = 0$ (contraction regime) through the logistic function:

$$p(\mathbf{x}) = \frac{1}{1 + e^{-\beta \cdot \mathbf{x}}}, \quad \beta \in \mathbb{R}^n,$$

and the objective function is the cross-entropy loss

$$\sum_{i=1}^m [-y^{(i)} \log(p(x^{(i)})) + (1 - y^{(i)}) \log(1 - p(x^{(i)}))].$$

Logistic regression can be implemented with ℓ_1 regularization, also known as LASSO regression, or ℓ_2 regularization, also known as Ridge regression, in order to prevent overfitting. These regularization methods add a penalty term including the ℓ_1 or ℓ_2 norm to the loss function in order to penalize overuse of parameters.

In this research, we select LASSO regression over other regularization methods because while other methods do not shrink the predictor variables' coefficients to 0, LASSO regression can shrink the coefficients to 0, ultimately performing both regularization and feature selection. Especially since we have 134 macroeconomic indicators as predictor variables when predicting the regimes, we aim to minimize the probability of overfitting by carrying out feature selection through ℓ_1 regularization.

LASSO regression, shortened for "Least Absolute Shrinkage and Selection Op-

erator,” is a regularized regression method that selects the predictors that have the greatest effect in predicting the response variable and shrinks other predictors in order to prevent overfitting (Tibshirani, 1996). The objective of ordinary logistic regression is

$$\sum_{i=1}^m \left[-y^{(i)} \log(p(x^{(i)})) + (1 - y^{(i)}) \log(1 - p(x^{(i)})) \right],$$

and the objective with regularization becomes

$$\sum_{i=1}^m \left[-y^{(i)} \log(p(x^{(i)})) + (1 - y^{(i)}) \log(1 - p(x^{(i)})) \right] + \lambda \sum_{k=1}^n |\beta_k|,$$

where $\lambda \in \mathbb{R}$ is a tuning parameter to control the effect of regularization.

4.2.2 Random Forests

Random forest, proposed by Breiman (2001), is an ensemble method of randomized tree predictors, used for the purpose of regression and classification. The trees used in random forests are grounded in the CART methodology, pioneered by Breiman et al. (1984).

CART decision tree is a binary recursive partitioning tree that partitions the predictor space based on the predictor variables. This method is capable of handling both classification and regression tasks, and is one of the most widely adopted non-parametric machine learning method.

Building on the principles of CART methodology, random forests use an ensemble approach of combining the outcomes of individual decision trees with random selection of features. This method amplifies the predictive accuracy and mitigates the risk of overfitting, a common limitation of decision trees. Consequently, random forests deliver more stable and reliable outcomes, surpassing the performance of individual tree-based models in various predictive modeling scenarios.

4.2.3 Neural Networks

In our research, we employ both the Feedforward Neural Network (FFNN) and Recurrent Neural Network (RNN) to discern the relationships between a diverse set of features and the regime information in commodity futures, enabling effective regime prediction based on these features.

Feedforward Neural Network (FFNN)

A Feedforward Neural Network (FFNN) is a machine learning method inspired by the human brain's architecture, and is capable of modeling the complex patterns in data through nonlinear transformations. It consists of interconnected nodes that perform computations, which form 3 layers organized in a hierarchical structure: input layers, hidden layers, and output layers. Neural networks learn the intricate patterns in data through a process of “learning”, by which it adjusts the weights of the connections between neurons. Although the process of learning is computationally expensive, the network processes new data rapidly and it can be generalized to diverse set of problems based on specific examples.

Given an FFNN with L layers with nonlinear activation functions g_l where $l \in [L]$ in each, we can write the output of a neural network for a generic input x , for hidden layers $1 \leq l \leq L$,

$$h^l = g^l(\mathbf{z}^l),$$

where

$$\mathbf{z}^l = \mathbf{W}_l^T h^l + \mathbf{b}^l$$

for $l \geq 2$, and

$$\mathbf{z}^1 = \mathbf{W}_1^T x + \mathbf{b}^1,$$

with learned parameters \mathbf{W}_l and \mathbf{b}^l . The final output \hat{y} will then become $g^L(\mathbf{z}^{L-1})$.

The commonly used nonlinear activation functions are the ReLU function, sigmoid function, or the hyperbolic tangent function.

Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are neural networks with recurrent connections that are capable of modeling the long-term dependencies of sequential or time series data. Since the dataset we use contains time series data, we experiment whether RNNs can also be a suitable option for extracting the regime patterns of the data.

Unlike FFNNs, RNNs maintain memory information through hidden states, and thus show a strength in processing a sequence of inputs. The network computes each hidden state h_t using the current input x_t and the previous hidden state \mathbf{h}_{t-1} as

$$\mathbf{h}_t = f(\mathbf{W}_x \mathbf{x}_t + \mathbf{W}_h \mathbf{h}_{t-1} + \mathbf{b}),$$

where f is the activation function, \mathbf{W}_x are the weights associated with the input to the states, \mathbf{W}_h are the weights associated with the previous hidden state, and \mathbf{b} is the bias term. Also, output at time t , y_t is computed as

$$\mathbf{y}_t = f(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}),$$

where \mathbf{W}_y are weights associated with the hidden layer connected to the output layer.

4.3 Portfolio Construction Methods

The following subsections illustrate the specific ways to construct a portfolio based on different strategies outlined in the literature review section: mean reversion, trend-following and risk parity.

4.3.1 Mean Reversion

Mean reversion strategies are based on the premise that asset prices and returns eventually move back towards their historical average. When the prices deviate from the average, investors can take it as a signal to either long or short positions based on the direction of the deviation. The most commonly used mean reversion signal is the asset price's deviation from the simple moving average (SMA). The strategy can be further optimized by adjusting weights based on the strength of the mean reversion signal or risk considerations.

4.3.2 Trend-Following

The goal of trend-following strategies is to understand the direction of asset price movements and make investment decisions that align with the trend. A commonly used indicator of the price trend is the exponential moving average (EMA), which is calculated as (Fong & Tai, 2009):

$$\text{EMA}_t = P_t - \text{EMA}_{(t-1)} \times \frac{2}{n+1} + \text{EMA}_{(t-1)}$$

where P_t is the price of the asset at time t , and n is the number of periods over which the average is calculated. These moving averages signal either up-trends or down-trends, and portfolio optimization involves reading these trends and making investment decisions accordingly to maximize returns.

4.3.3 Risk-Parity

Risk-parity strategies allocate capital across portfolio assets so that each asset has equal risk contribution. The optimal weights for each asset could be derived by solving the following optimization problem (Maillard et al., 2010),

$$\begin{aligned}
& \min_w \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} \\
\text{s.t. } & \sum_i b_i \ln w_i \geq c \\
& \mathbf{1}^T \mathbf{w} = 1 \\
& \mathbf{w} \geq 0,
\end{aligned}$$

where w_i is the weight of asset i , b_i is the relative risk contribution of asset i , Σ is the covariance matrix of all assets, and c is a fixed positive constant.

4.3.4 Regime Switching Framework

Mean Reversion

The mean reversion strategy is adapted with our regime-switching framework by introducing regime probabilities. We use the predicted regime probabilities from the prediction models to modulate the signals generated by the mean reversion strategy's trading rule. For instance, even if the strategy indicates a strong buy strategy, if the predictive model anticipates a contraction regime, we moderate the signal. The degree of modulation corresponding to the predicted probability, ultimately serving as the model's confidence in such regime change being realized.

Trend Following

Using the regime prediction model trained on the primary commodity index data and macroeconomic indicators, we predict regime shifts and generate trading signals accordingly. Although we adhere to the trend-following strategy's general rule by using the moving averages to extract the up-trends and down-trends, and ultimately the buy or sell signals, the intensity of the signals is further adjusted based on the probability of a predicted regime.

Portfolio weights are dynamically adjusted based on the generated signals and the current market regime. The portfolio rebalances on a predetermined frequency or as determined necessary based on significant changes in the market regime.

Risk-Parity

We modify the traditional risk-parity portfolio described in Section 4.3.3 to dynamically adjust the portfolio based on the prevailing regime. We formulate the optimization problem as described in Section 4.3.3, but estimate the covariance matrix Σ differently by accounting for regime probabilities. The regime-switching risk-parity portfolio uses a weighted covariance, a covariance matrix computed based on the regime probabilities obtained from the the regime prediction model. The weighted covariance matrix is computed as follows:

$$\hat{\Sigma}_{\text{weighted}} = P_{\text{growth}} \cdot \hat{\Sigma}_{\text{growth}} + P_{\text{contraction}} \cdot \hat{\Sigma}_{\text{contraction}},$$

where P_{growth} and $P_{\text{contraction}}$ are the probabilities of growth and contraction regimes respectively, and $\hat{\Sigma}_{\text{growth}}$ and $\hat{\Sigma}_{\text{contraction}}$ are the covariance matrices for those regimes.

The portfolio is rebalanced based on a predetermined period, with assets being reallocated according to the optimized weights of the portfolio.

4.3.5 Backtesting Framework

For backtesting, we construct four distinct portfolios with varying degrees of diversification to test the efficacy of our regime-switching framework. The composition of these portfolios is chosen to explore various asset combinations and their performance under different market conditions, thus providing a robust framework for backtesting. See Table 4.1 for the specific composition of each portfolio. Also, see Figure 4.1 for the plot of cumulative log return values of the assets that form the different portfolios.

	Components
Portfolio 1	GSCI(DBC), SPY, TLT
Portfolio 2	GSCI(DBC), SPY, TLT, SBBMGLU
Portfolio 3	GSCI(DBC), SPY, TLT, IEF, SHY, XSVM
Portfolio 4	GSCI(DBC), SPY, TLT, IEF, SHY, XSVM, SBBMGLU, IGSB

Table 4.1: Components of Each Portfolio

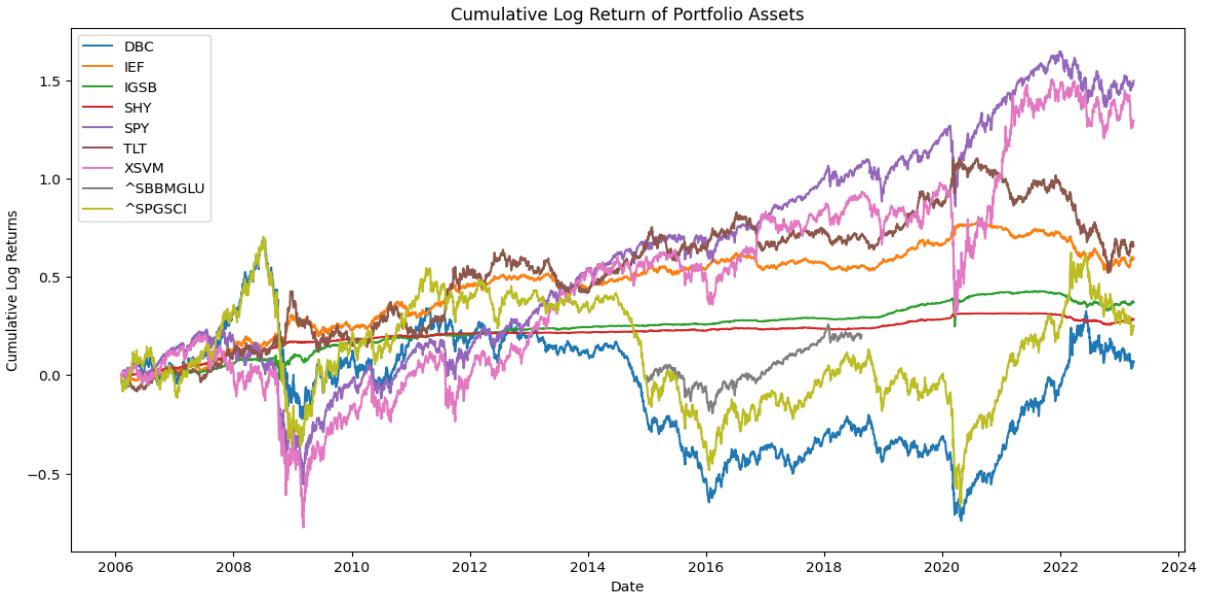


Figure 4.1: Comparison of Cumulative Returns for Portfolio Assets

Portfolio 1 consists of a commodity futures index, either the GSCI or DBC Index, the SPDR SP 500 ETF (SPY), and the iShares 20+ Year Treasury Bond ETF (TLT), encapsulating a traditional mix of commodities, equities, and long-term government bonds. Portfolio 2 extends Portfolio 1 by adding the S&P Global Broad Market Index (SBBMGLU), a global equity index that includes more than 14,000 stocks from developed and emerging markets, for further diversification. By including global equities, we attempt to understand the compatibility of our regime-switching approach with the varying growth rates and exposures to different currencies and economic cycles.

Portfolio 3 further diversifies the asset mix by including the iShares 7–10 Year Treasury Bond ETF (IEF), the iShares 1–3 Year Treasury Bond ETF (SHY), and the Invesco SP SmallCap Value with Momentum ETF (XSVM). This portfolio incorpo-

rates a wider range of fixed income securities with varying maturities and a small-cap value equities as well. Portfolio 4, the most diversified of the group, includes all the components of Portfolios 1 through 3, with the addition of the iShares Short-Term Corporate Bond ETF (IGSB), further enriching the portfolio's fixed income spectrum. Portfolio 4 showcases a portfolio that spans small-cap to large-cap equities, domestic to global equities, and a mix of corporate and government bonds, further with the inclusion of the commodity index, aimed at providing a nuanced understanding of how different asset allocations respond to regime changes.

To evaluate the performance of the portfolios, we utilize several key metrics: annualized returns (Ret.), annualized volatility (Vol.), and maximum drawdown (Max DD). Additionally, we incorporate three performance metrics — the Sharpe, Calmar, and Sortino Ratio — each of which provides insights into the risk-adjusted returns of the portfolios. Then, we compare these metrics against those of two traditional benchmark portfolios: the Equal Weight portfolio and the 60/40 portfolio. The three ratios are further detailed in the subsequent subsections. Note that we assume a zero risk-free rate, $r_f = 0$ when calculating the three ratios.

Sharpe Ratio

The Sharpe Ratio is a widely used risk-adjusted return measure that calculates the excess return per unit of deviation in an investment. Specifically, the Sharpe Ratio is defined as the following equation:

$$\text{Sharpe Ratio} = \frac{\mu_p - r_f}{\sigma_p},$$

where r_f is the risk-free rate (or a benchmark return), μ_p is the average return, and σ_p is the standard deviation of the portfolio's excess returns. The Sharpe Ratio allows for a comparative analysis of investments through integration of both the reward and

risk components. A higher Sharpe Ratio indicates a more attractive risk-adjusted return, suggesting that the investment generates a greater reward per unit of risk.

Calmar Ratio

The Calmar Ratio serves as another risk-adjusted performance metric, focusing on the relationship between the annualized rate of return and the maximum drawdown during a specified time frame. The Calmar Ratio is calculated as the following:

$$\text{Calmar Ratio} = \frac{\mu_p - r_f}{\text{Maximum Drawdown}},$$

The maximum drawdown is a measure of the largest peak-to-trough decline in the investment value. A higher ratio implies that the return of the investment was significant enough compared to the risk of significant drawdowns, ultimately indicating that the investment has a higher returns on a risk-adjusted basis over the timeframe.

Sortino Ratio

The Sortino Ratio is a variation of the Sharpe ratio that focuses solely on downside risk, rather than overall volatility. This ratio differentiates harmful volatility from total volatility to provide a more nuanced assessment of investment risk. The Sortino Ratio is calculated as the following:

$$\text{Sortino Ratio} = \frac{\mu_p - r_f}{\sigma_d},$$

where σ_d is the downside deviation. This approach recognizes that investors are primarily concerned with the risk of losses, rather than the fluctuations associated with above-average returns. A higher Sortino Ratio indicates that the investment yields more return per unit of bad risk taken. The Sortino Ratio is particularly valuable for evaluating portfolios with non-normal return distributions, providing a

more accurate depiction of risk-adjusted performance by penalizing only the volatility of negative returns.

Chapter 5

Results and Analysis

5.1 Regime Identification Outcomes

In this section, we depict the regime identification outcomes yielded by the trend-filtering algorithm and the continuous statistical jump model. We include the implementation details and results for both the GSCI and DBC index.

5.1.1 Trend Filtering

The following are the outcomes for regime identification when we use the ℓ_1 trend-filtering algorithm. See Figure 5.1 for the plot of trend components extracted from the log-returns through ℓ_1 trend filtering with varying values of λ , in comparison with the observed values. We notice the trend that as the λ value increases, the number of piece-wise linear components decrease due to the penalty in the objective function. Accordingly, the identified number of regime shifts decreases.

We first aim to find the λ value that maximizes the difference in the next period's returns from the predefined range of λ values. We conduct time series cross-validation to ensure that the optimal λ value that we obtain is a robust outcome. After validation, we discover that $\lambda = 0.08$ has the greatest difference. See Figure 5.1 and Figure

5.2 for detailed plots illustrating such outcomes. The green highlighted regions in the plots indicate the contraction regimes identified through the algorithm.

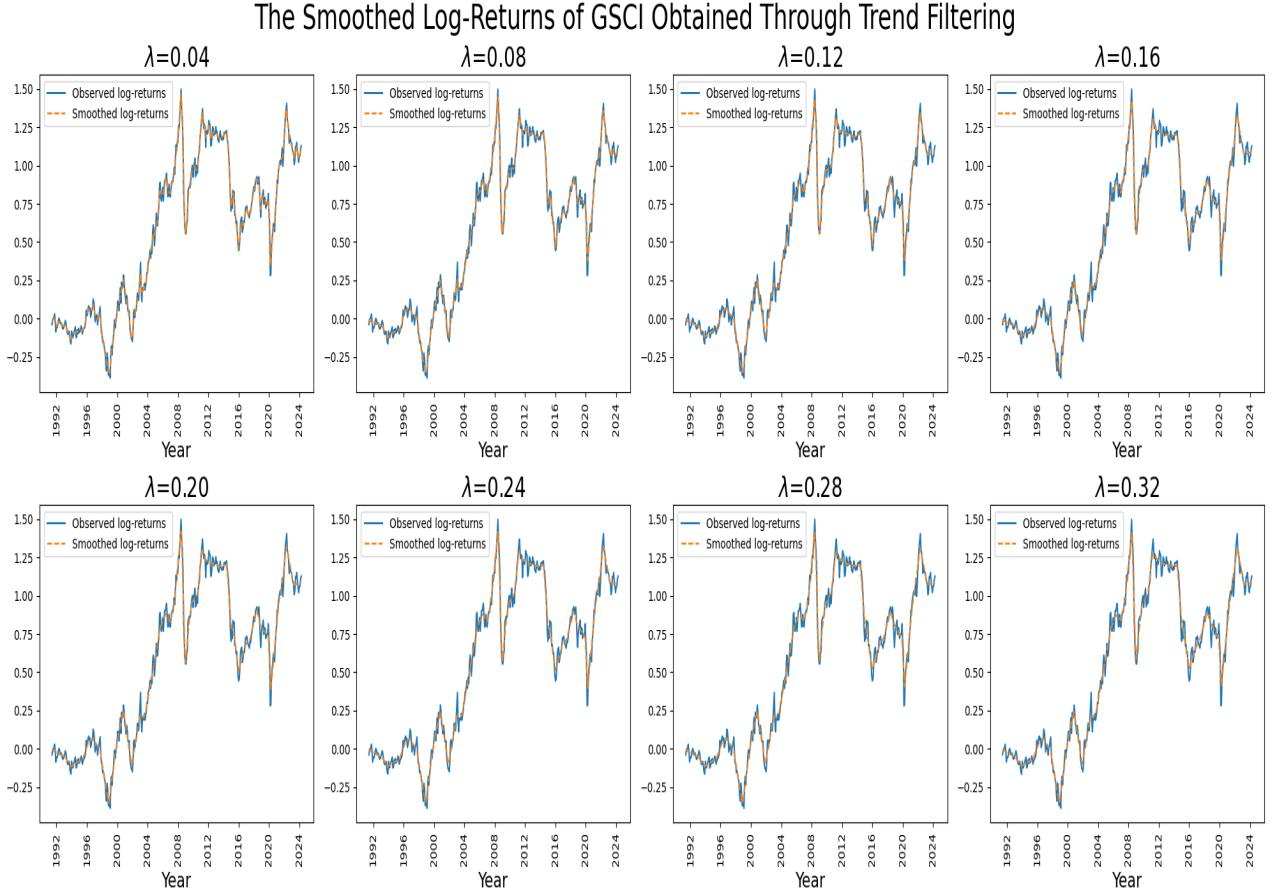


Figure 5.1: Comparison of Trend-filtering with Different Hyperparameter λ

The outcomes of ℓ -1 trend filtering heavily depend on the selection of the tuning parameter λ . The approach of finding the parameter λ that maximizes the difference in the next period's return seems to yield an output that is discontiguous. Thus, we implement Yamada (2018)'s approach for hyperparameter selection.

Yamada (2018) proposes a methodology of tuning the λ parameter for ℓ -1 trend filtering algorithm based on the fact that there is more knowledge on how to select the tuning parameter for HP filtering. Through mathematical proof, they formulate

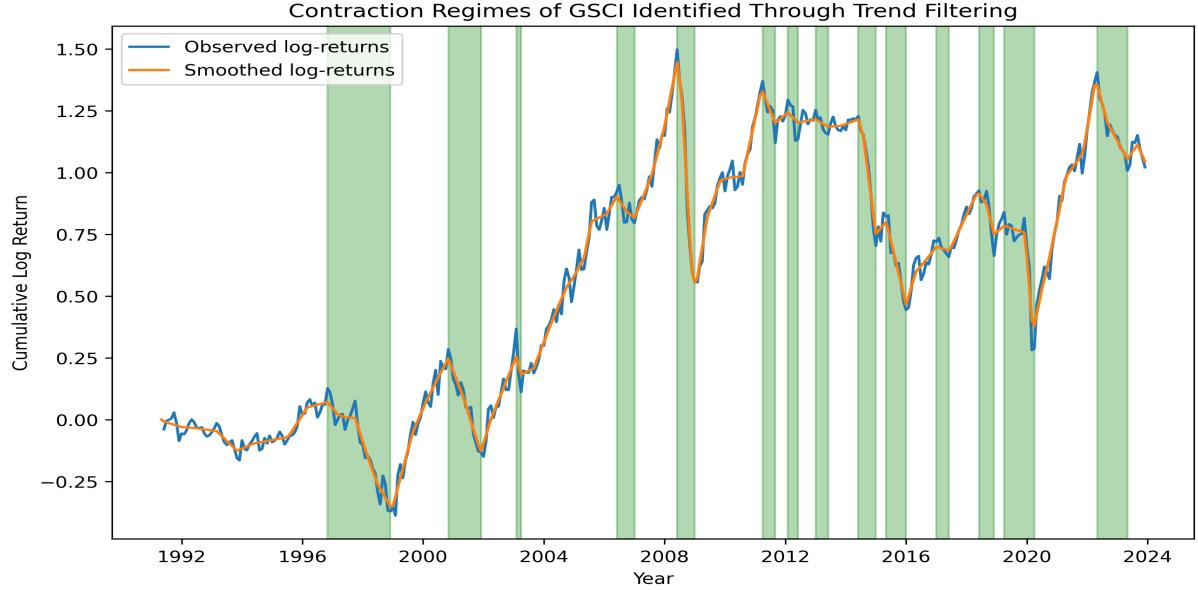


Figure 5.2: Contraction Regimes Identified Through Trend Filtering with the Tuned λ Value

the problem as the following constrained minimization problem

$$\hat{\mathbf{x}}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{D}\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - \mathbf{x}\|_2^2 \leq c, \hat{\mathbf{u}}^* = \mathbf{y} - \hat{\mathbf{x}}^*,$$

where $\mathbf{D} \in \mathbb{R}^{(n-2) \times n}$ is a second-order difference matrix, and

$$c = \|\hat{\mathbf{u}}^{\text{hp}}\|_2^2 = \|\hat{\mathbf{u}}^{\text{lt}}\|_2^2 = \|\hat{\mathbf{u}}^*\|_2^2.$$

As a result,

$$\lambda = 2\|(\mathbf{DD}')^{-1}\mathbf{D}\hat{\mathbf{u}}^*\|_\infty.$$

See Figure 5.3 for the regime identification outcome with hyperparameter tuning.

We observe that the implementation of Yamada's approach for hyperparameter tuning yields a more contiguous outcome, compared to the outcome without the tuning process, with even slight drops of returns being identified as a contraction regime.

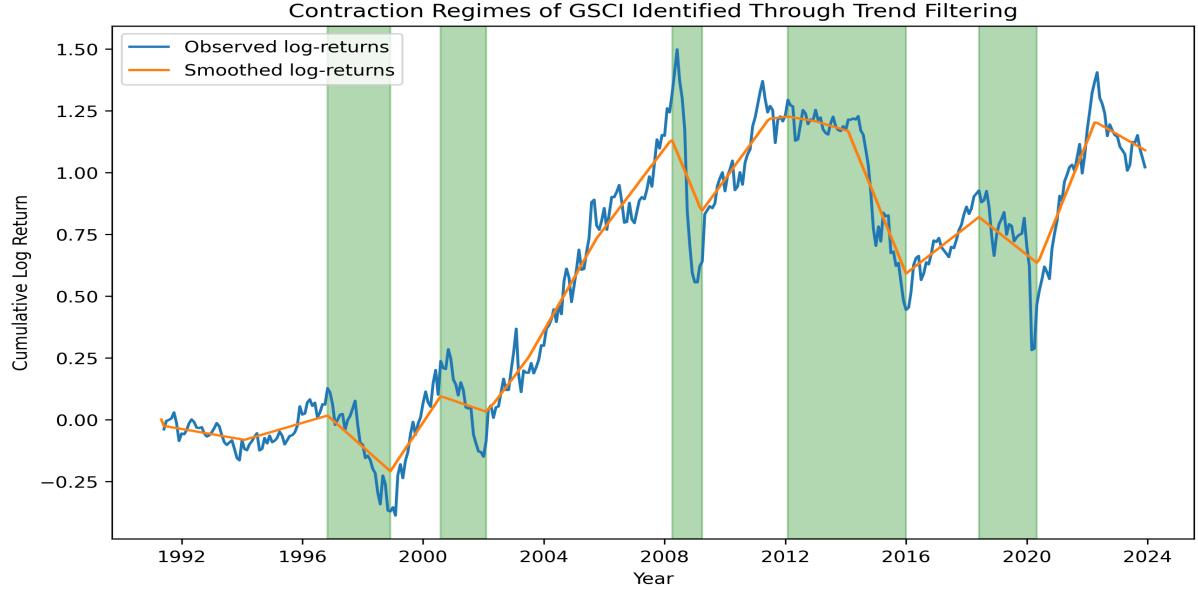


Figure 5.3: Contraction Regimes Identified in GSCI Data Through Trend Filtering with Yamada's Hyperparameter Tuning

We repeat the same steps for the DBC index. See Figure 5.4 and Figure 5.5 for the corresponding plots. Before hyperparameter tuning, $\lambda = 0.04$ yields the greatest difference in the next periods' returns (differences in returns for the identified regimes) after time series cross-validation.

Yamada's hyperparameter tuning approach in regime identification yields more stable results than methods without tuning, which appear sensitive to minor changes in returns. While implementing trend-filtering with hyperparameter tuning is more promising, it is noteworthy that the increased penalty for regime switches can oversimplify time series data. This result highlights a trade-off between stability and data complexity in regime identification.

These findings depict the high variability still present in the regime identification outcomes, influenced by the data and the choice of the hyperparameter λ . Also, the commodity markets' regimes are not perfectly aligned with the overall economic cycles identified by the National Bureau of Economic Research unlike stock markets, posing a challenge for direct comparison. These outcomes point towards the possibility of using

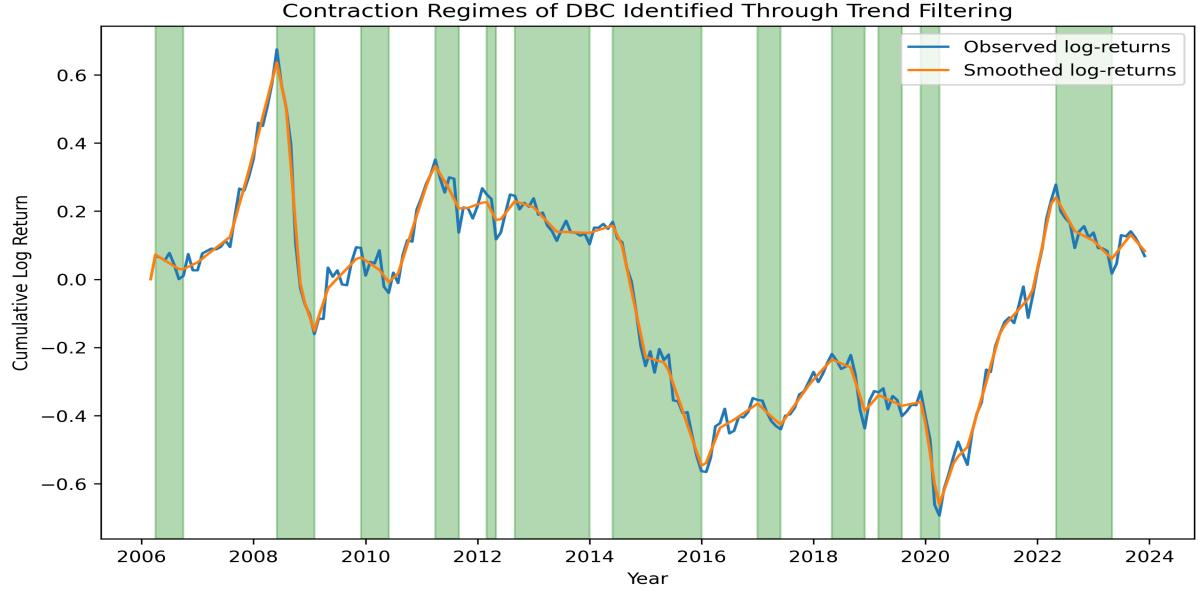


Figure 5.4: Contraction Regimes Identified in DBC Data Through Trend Filtering with the Tuned λ Value

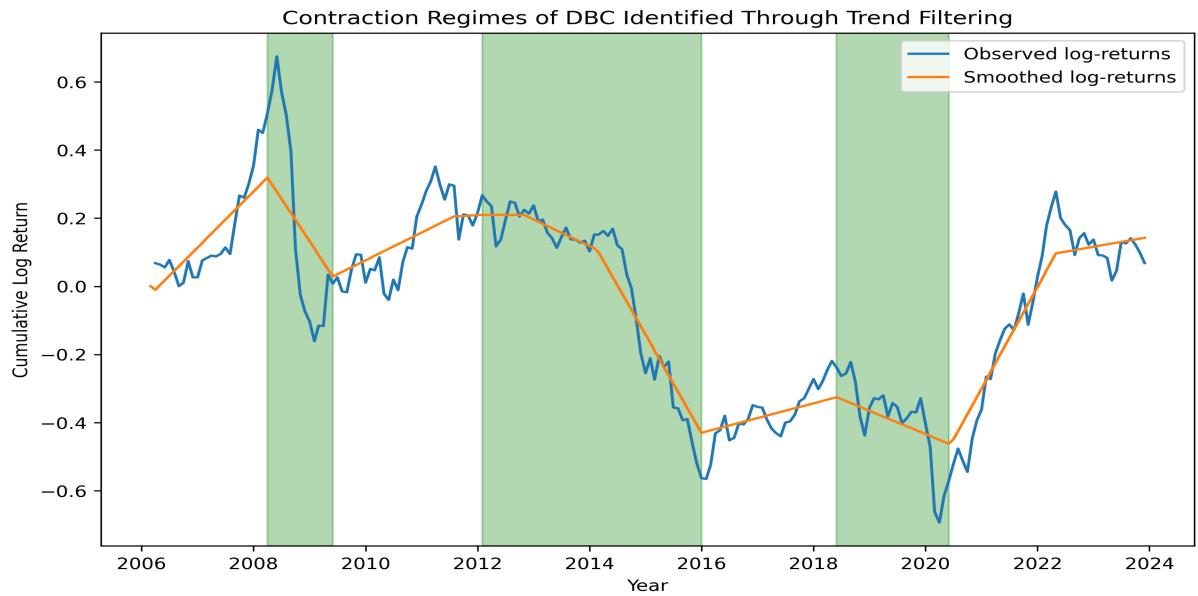


Figure 5.5: Contraction Regimes Identified in DBC Data Through Trend Filtering with Yamada's Hyperparameter Tuning

a probabilistic approach for a nuanced understanding of the regimes latent in the time series data. Continuous statistical jump models enable such probabilistic information since these models facilitate the probabilistic order-dependent clustering of different

states. Thus, we move on to implementing jump models for regime identification, with the ultimate goal of being able to stabilize the identification of regimes within both the DBC and GSCI indices' data, and improve the quality of regime prediction based on these identifications.

5.1.2 Continuous Statistical Jump Models

This subsection presents the outcomes of regime identification from the application of continuous statistical jump models, as proposed by Aydinhan et al. (2023). We implement the model using the coordinate descent algorithm, using 80% of the dataset for training and the other 20 percent of the data for testing. In order to select the best jump penalty parameter λ , we conduct time series cross-validation for both datasets. As a result, we learn that the best λ is $\lambda = 0.05$ for the GSCI data and $\lambda = 0.03$ for the DBC data.

Since regime identification is a form of unsupervised learning applied to unlabeled time series data, we devise an evaluation method of using the jump model's total loss that incorporates both the difference between the trend series and the observation data, and the jump penalty for cross-validation and model evaluation. See Table 5.1 for specific loss values when conducting hyperparameter tuning on the GSCI training dataset. The final model performance for $\lambda = 0.05$ on the test data is 0.3143. Also, see Figure 5.6 and Figure 5.7 for the plot of regime identification and the probabilities of contraction regimes in each time period, respectively. Again, the green highlighted regions illustrate the contraction regimes identified through the model.

The identified regimes appear consistent with the identification outcomes from the ℓ_1 trend-filtering algorithm in general, but seem to be more contiguous, which we attribute to the jump penalty factor which prevents the model from being overly sensitive to market fluctuations.

We apply the same methodology to the DBC dataset and observe the outcomes as

λ	Loss
0.01	0.5205
0.02	0.2448
0.03	0.5015
0.04	0.4799
0.05	0.2332
0.06	0.5354
0.07	0.2470
0.08	0.4917

Table 5.1: GSCI Jump Model Loss for Different λ Values

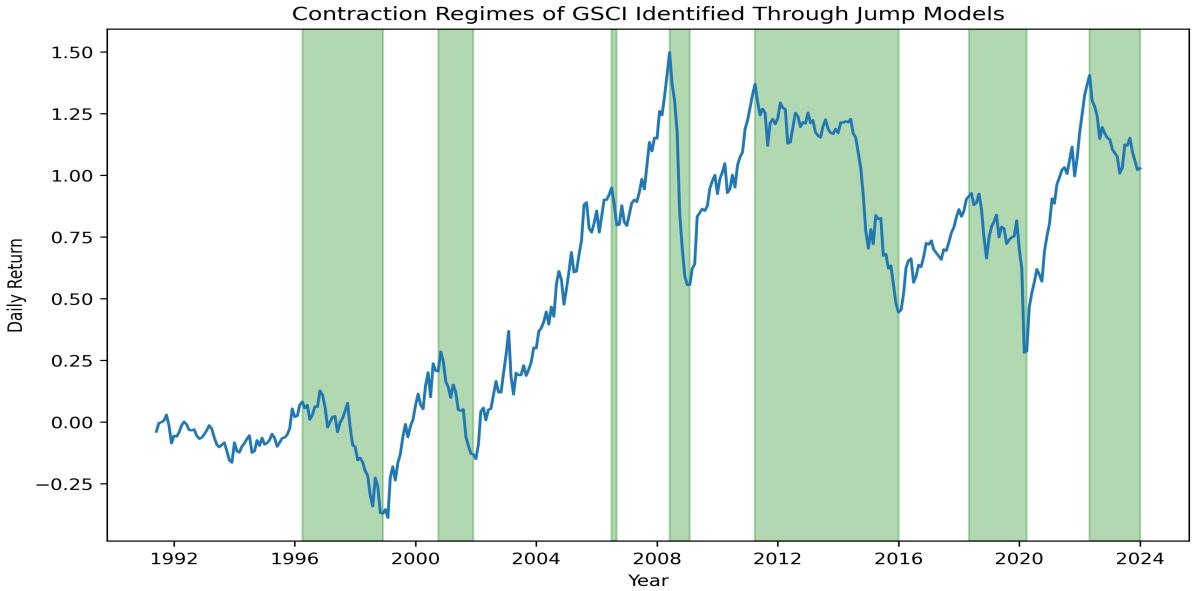


Figure 5.6: Contraction Regimes Identified in GSCI Data Through Jump Models

illustrated in Table 5.2. Also, we obtain the equivalent plots in Figure 5.8 and Figure 5.9 for the regime identification outcomes and probabilities, respectively. The final model performance for $\lambda = 0.03$ on the test data is 0.3176. Although the results seem to mostly align with the outcomes of ℓ_1 trend filtering, one noteworthy difference is that the jump model detects the period between 2022 and 2024 as a contraction regime, unlike the trend-filtering algorithm. This observation is interesting considering that although jump models are intended to penalize shifts in regimes through its jump penalty hyperparameter, in some occasions they might show greater sensitivity to regime changes. The time series data for the corresponding period seems to display

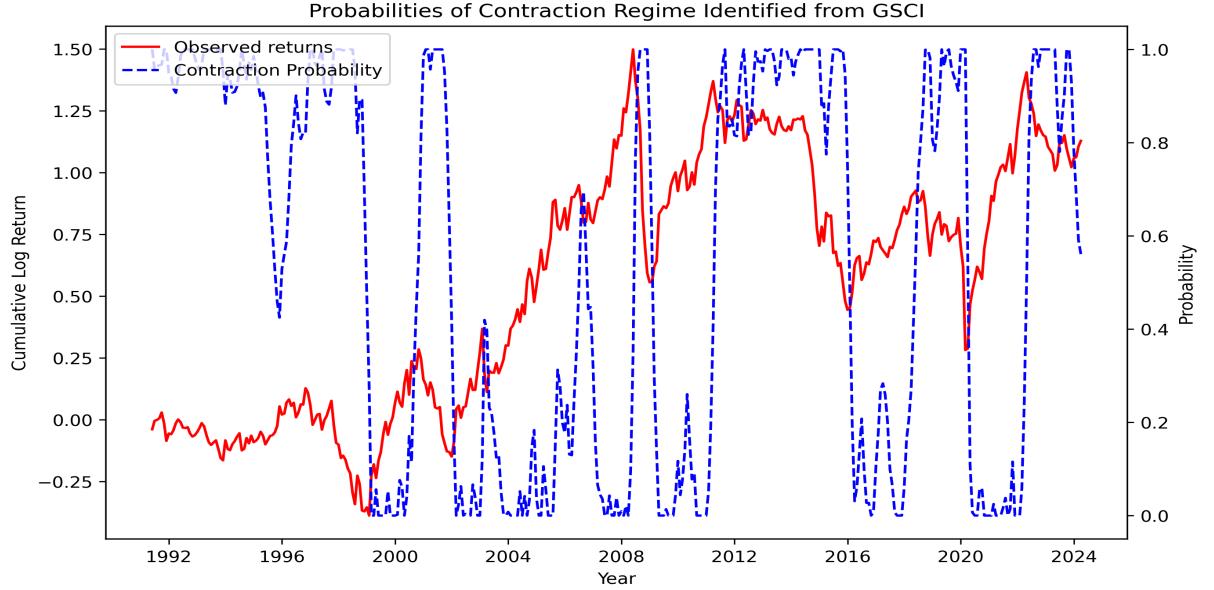


Figure 5.7: Probabilities of Contraction Regime in GSCI Data

trends of contraction regimes, so we conclude that the jump model outperforms the trend-filtering algorithm for precise identification of regimes in the DBC data.

λ	Loss
0.01	0.1812
0.02	0.1910
0.03	0.0950
0.04	0.2032
0.05	0.1456
0.06	0.0951
0.07	0.1058
0.08	0.2116

Table 5.2: DBC Jump Model Loss for Different λ Values

Overall, the continuous statistical jump model appears to yield more promising outcomes for contraction regime detection. Also, the model has a significant advantage of generating a probability distribution. This feature is particularly valuable, offering information that could be leveraged for the adoption of a probabilistic regime-switching approach. After the comparison of these two approaches for regime identification, we move on to using the obtained regime labels for further regime prediction

and portfolio construction.

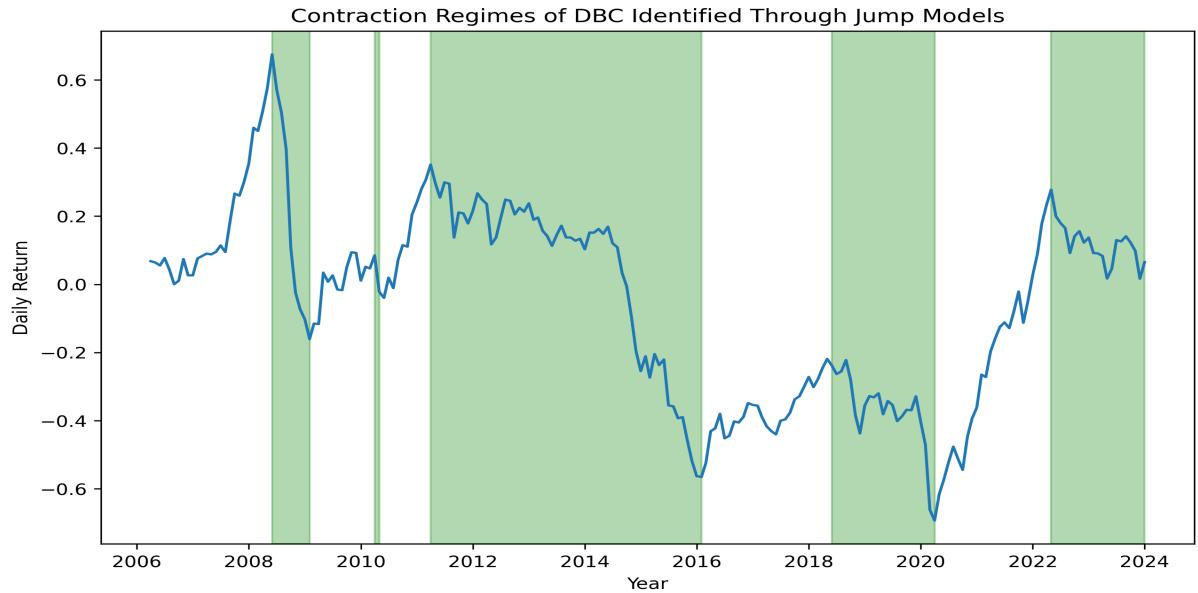


Figure 5.8: Contraction Regimes Identified in DBC Data Through Jump Models

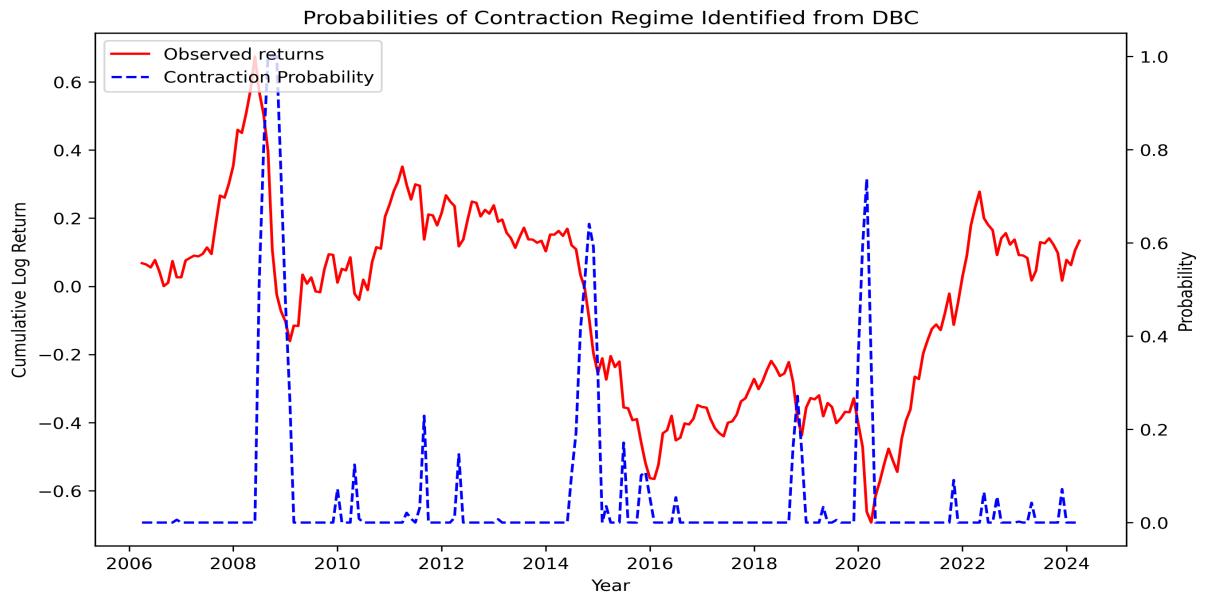


Figure 5.9: Probabilities of Contraction Regime in DBC Data

5.2 Regime Prediction Outcomes with Trend Filtering

As mentioned in Section 4.2, we first use the discrete regime labels obtained by the ℓ_1 trend filtering algorithm and train different machine-learning models — logistic regression, recurrent neural networks, feedforward neural networks, random forests and gradient boosting — with the monthly returns along with the 134 macroeconomic variables obtained from the FRED-MD database. In order to prevent look-ahead bias, we preserve temporal order when splitting the data into the training and testing set. Moreover, we adopt an approach where we incrementally make predictions on the test data with an expanding window. For instance, when predicting the regime shifts that occur in GSCI data, we start with 40 percent of training data and use an expanding window of 8 months to make predictions on the test data. The sizes of the training data and expanding window are determined through empirical testing. Note that the plots in this section span the period that is included in the test data, unlike the previous section in which the plots span the entire time period starting from the indices' inception dates. See Table 5.3 for the regime prediction accuracy on GSCI data.

Model	Test Accuracy (%)
ℓ_1 Logistic Regression	90.21
Gradient Boosting	88.94
FFNN	72.77
RNN	72.77
Random Forest	91.06

Table 5.3: GSCI Regime Prediction Test Accuracy with Regime Information Obtained Through Trend Filtering

Considering that the proportion of contraction regimes in the dataset is 44.76%, we conclude that the models' prediction perform significantly better than mere guessing.

We proceed to make predictions on the DBC index data. In light of the fact that

there is less data available for the DBC index due to its later inception date than that of GSCI, we empirically set the size of the expanding window to be 6 months. We use the same set of models for prediction. See Table 5.4 for the test accuracy yielded by the different prediction models.

Model	Test Accuracy (%)
ℓ_1 Logistic Regression	90.63
Gradient Boosting	91.41
FFNN	46.88
RNN	36.72
Random Forest	89.84

Table 5.4: DBC Regime Prediction Test Accuracy with Regime Information Obtained Through Trend Filtering

The following subsections provide an analysis of the outcomes of each of the models. The plots included outline the predictions made on the test dataset.

5.2.1 Logistic Regression

When conducting regime prediction on the GSCI data, the ℓ_1 logistic regression yields a test accuracy comparable to that of random forests, ranking second in test accuracy among the five models. At the same time, the model shrinks 99 predictor variables to 0, effectively conducting the role of feature selection to prevent overfitting. The predicted contraction regimes display some discontinuity as shown in Figure 5.10, which we believe is caused by the noise in the regime labels as well as a consequence of using regularization.

ℓ_1 logistic regression also yields the second-highest test accuracy in predicting regimes on the DBC dataset. The model accurately and consistently predicts most regimes, including certain periods that are not flagged by the trend-filtering algorithm as contraction regimes, as depicted in Figure 5.11. These periods in fact show down-trends in the cumulative log returns, suggesting that the under-identification might stem from the hyperparameter tuning during the regime identification process rather

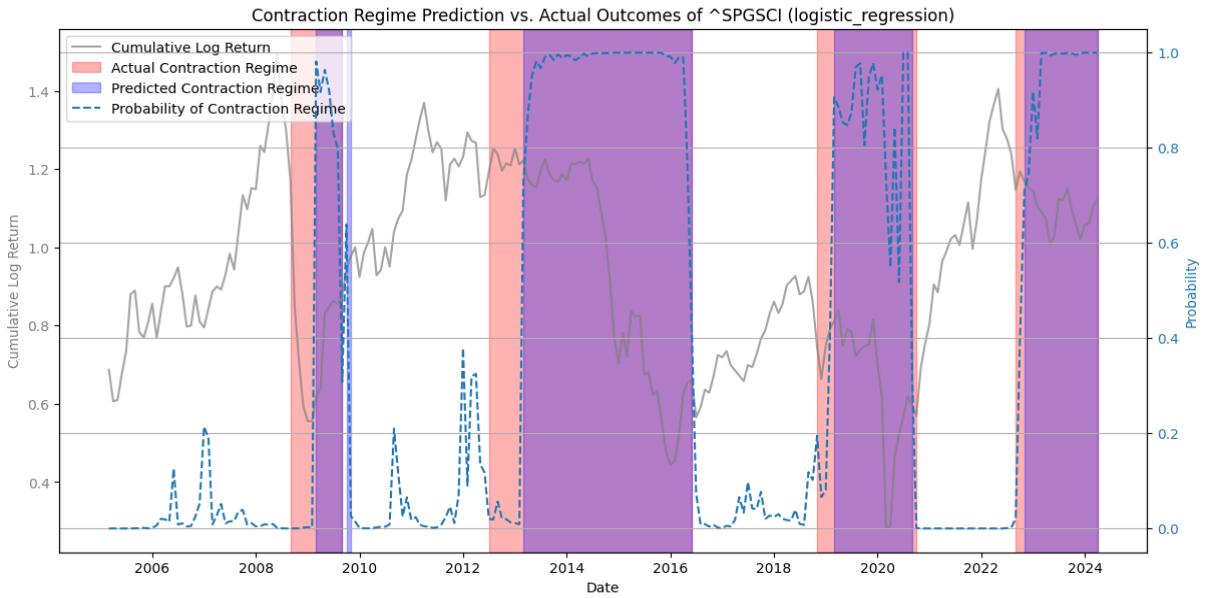


Figure 5.10: Contraction Regimes of GSCI Predicted Through Trend Filtering + Logistic Regression

than a deficiency of the predictive model itself. However, the model also predicts brief and scattered periods that do not appear to be a contraction regime, indicating opportunities to enhance the precision of predictions.

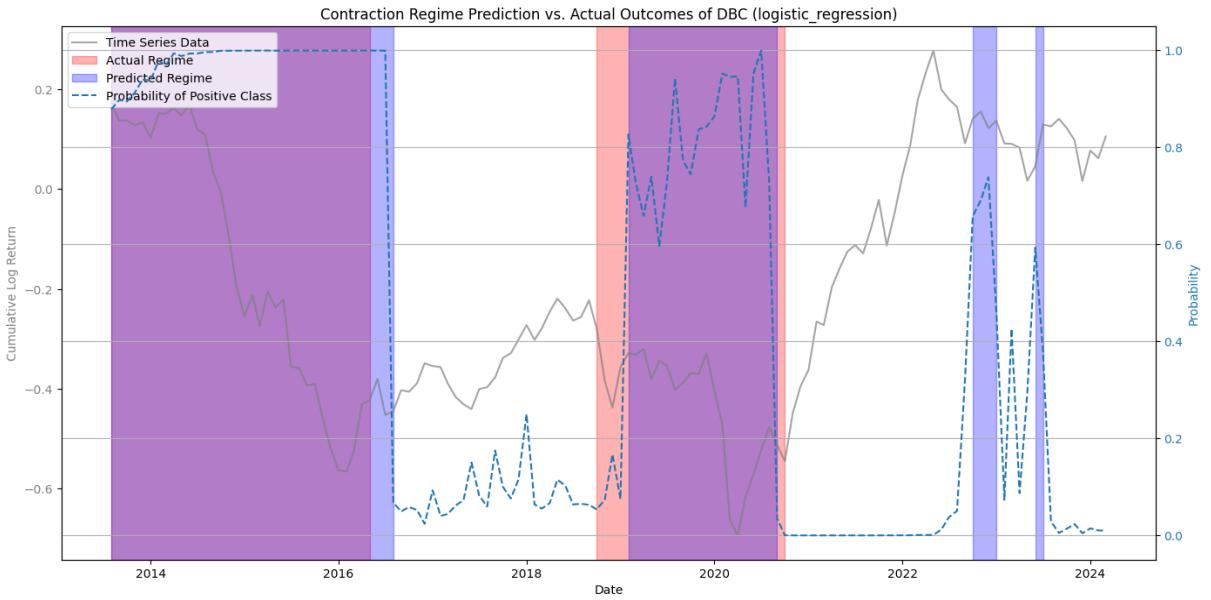


Figure 5.11: Contraction Regimes of DBC Predicted Through Trend Filtering + Logistic Regression

5.2.2 Neural Networks

Both the Feedforward Neural Network (FFNN) and the Recurrent Neural Network (RNN) yield poor levels of test accuracy for both GSCI and DBC data, with test accuracies being as low as 72.77% for GSCI and 36.72% for DBC. As shown in Figure 5.12 and Figure 5.13, the majority of predictions made using the FFNN seem to be inconsistent with the regime labels, yielding a lot of false negatives and false positives.

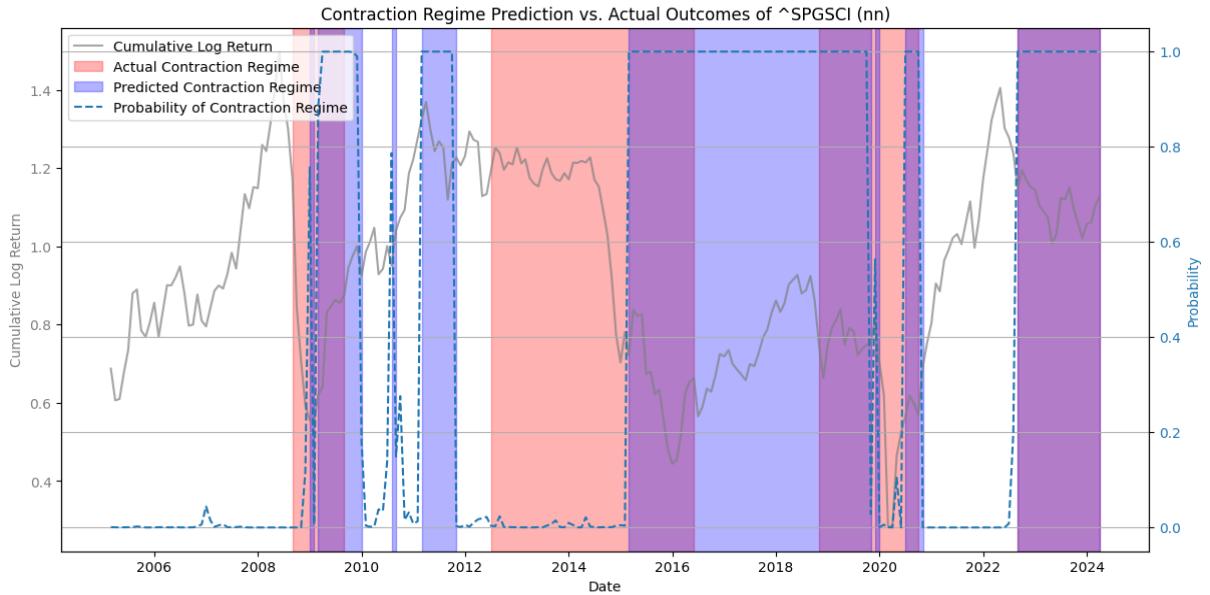


Figure 5.12: Contraction Regimes of GSCI Predicted Through Trend Filtering + Feedforward Neural Networks

Although RNNs are known for their strength in modeling sequential data, interestingly their performance is very poor, as shown in Figure 5.14 and Figure 5.15. Unlike the FFNN, most of the incorrect predictions are centered around false positives.

Overall, the performances of both neural networks are disappointing, so we decide against using neural networks for the prediction task.

5.2.3 Random Forests

Random forests yield the highest test accuracy in regime prediction for GSCI. Through *Scikit-learn's GridsearchCV* function along with time series cross-validation for hy-

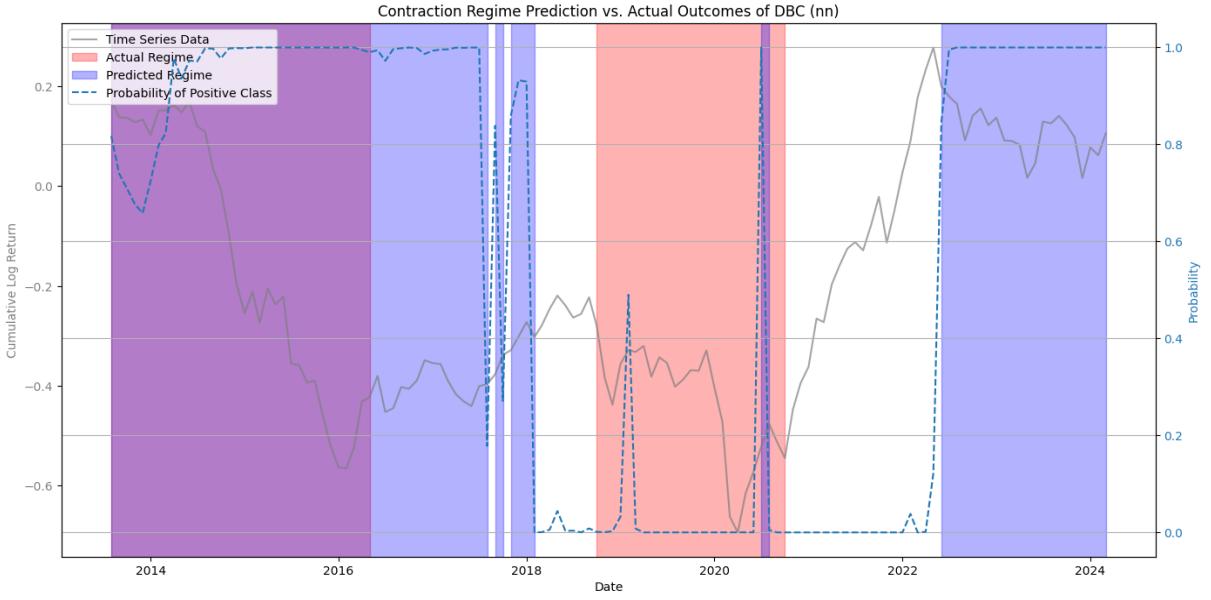


Figure 5.13: Contraction Regimes of DBC Predicted Through Trend Filtering + Feedforward Neural Networks

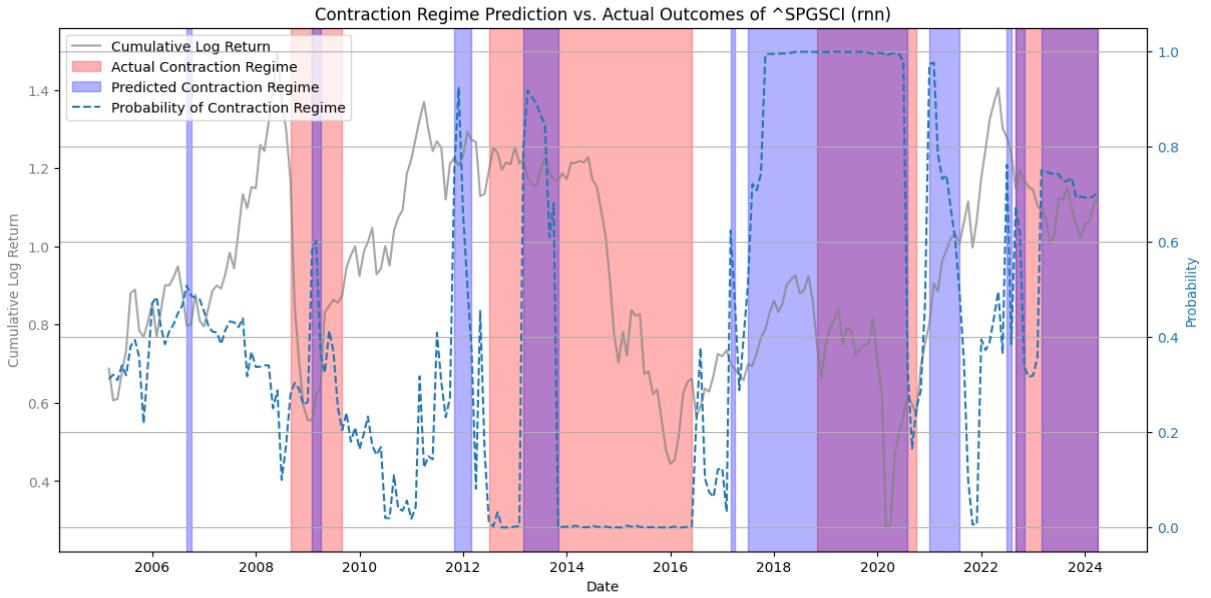


Figure 5.14: Contraction Regimes of GSCI Predicted Through Trend Filtering + Recurrent Neural Networks

After parameter tuning, we enhance the performance up to a test accuracy of 92.31%. See 5.16 for a detailed illustration of the predictions. The predictions seem to mostly align with the regime labels obtained through the application of trend-filtering.

When predicting regimes from the DBC time series data, random forests are

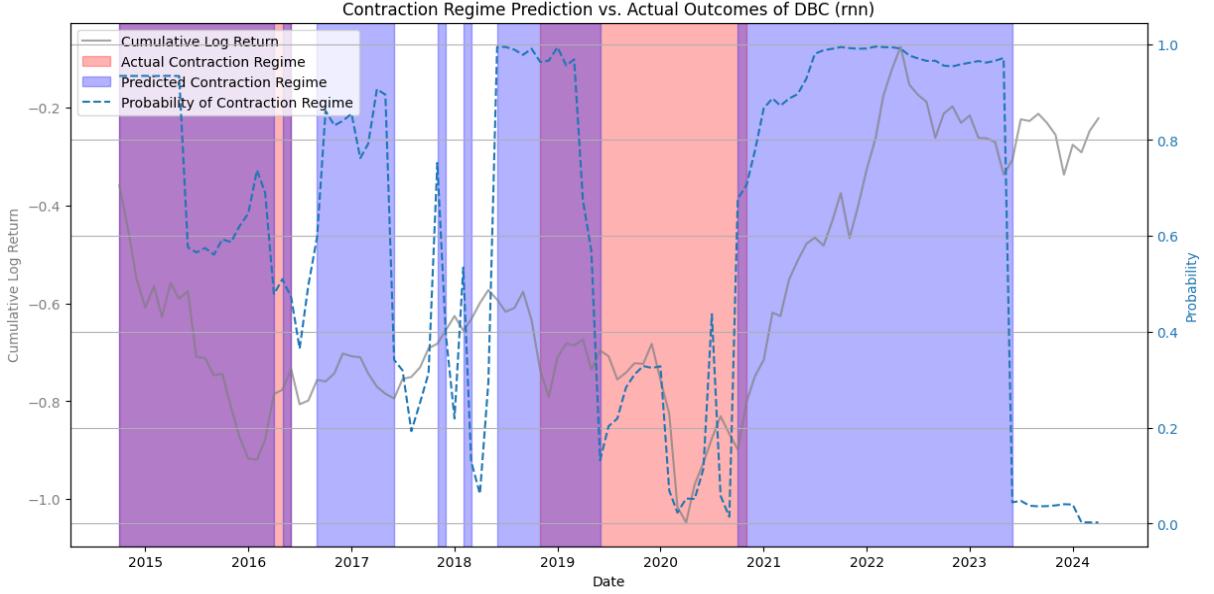


Figure 5.15: Contraction Regimes of DBC Predicted Through Trend Filtering + Recurrent Neural Networks

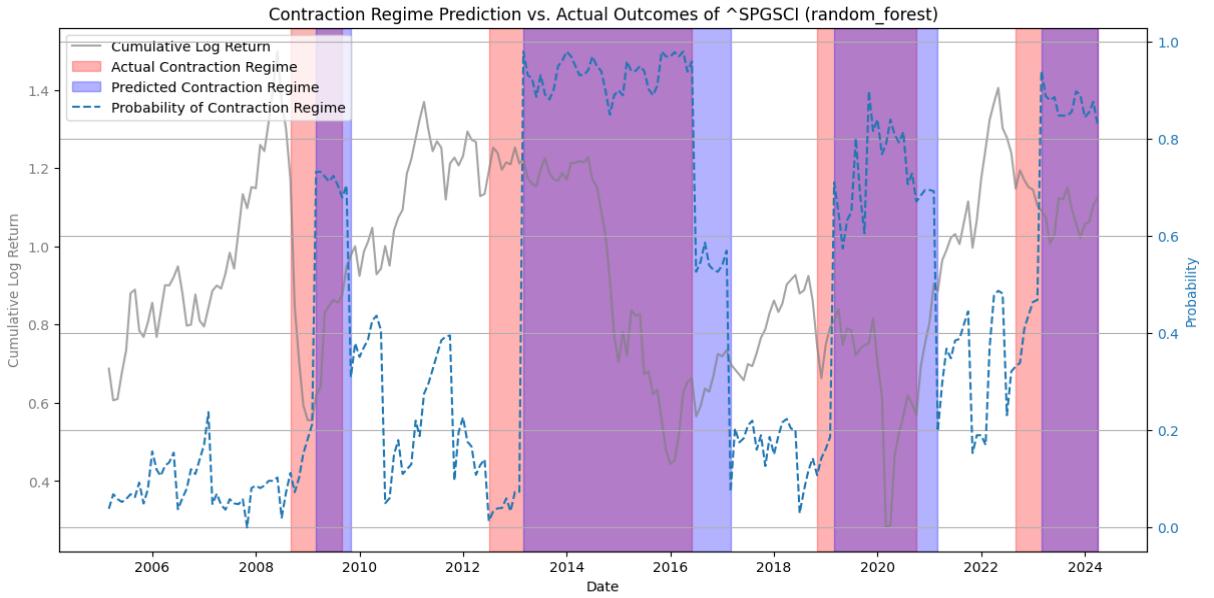


Figure 5.16: Contraction Regimes of GSCI Predicted Through Trend Filtering + Random Forests

outperformed by ℓ_1 logistic regression and gradient boosting. However, as illustrated in Figure 5.17, the predicted outcomes seem to be contiguous and mostly overlap with the highlighted regions labeled through trend-filtering.

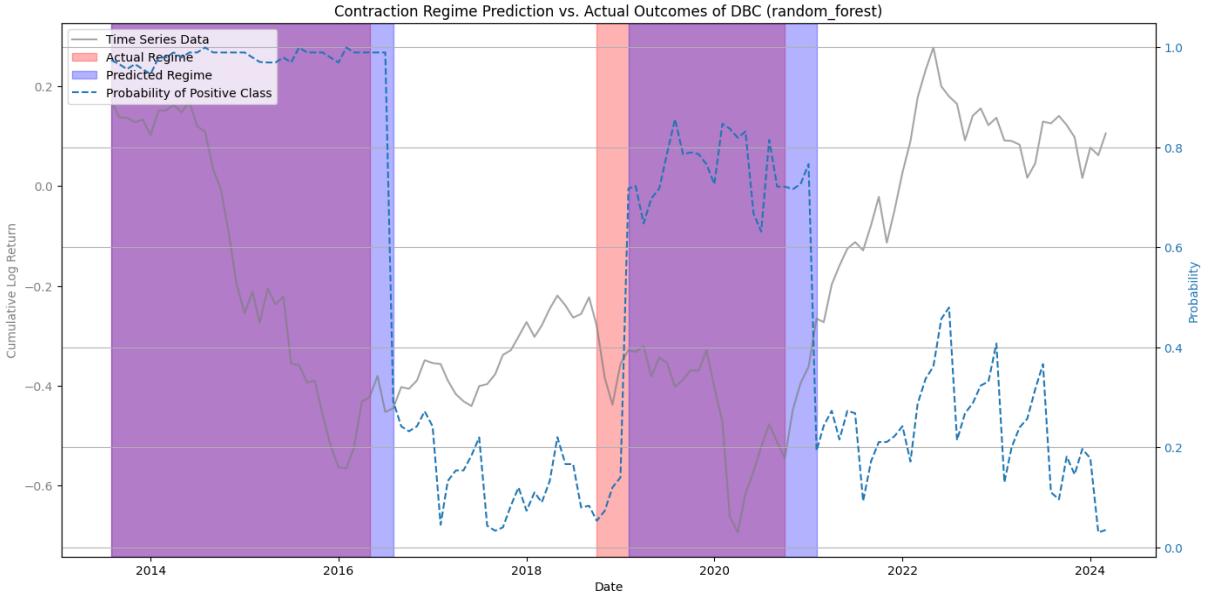


Figure 5.17: Contraction Regimes of DBC Predicted Through Trend Filtering + Random Forests

5.2.4 Gradient Boosting

The predictions made using gradient boosting are mostly accurate, but the predictions made on the GSCI data seems to be more discontiguous compared to the predictions made with random forests. The outcomes are clearly more promising than those from neural networks, but are outperformed by logistic regression and random forests. Meanwhile, gradient boosting yields the highest test accuracy for the DBC dataset, suggesting that the same model could yield very different performances based on the nature of the dataset. See Figure 5.18 and Figure 5.19 for the prediction outcomes.

5.3 Regime Prediction Outcomes with Statistical Jump Models

We move on to using the regime information obtained by jump models and again train the five machine learning models, using the same set of macroeconomic variables and commodity index data, as well as the same methodology of time series cross-

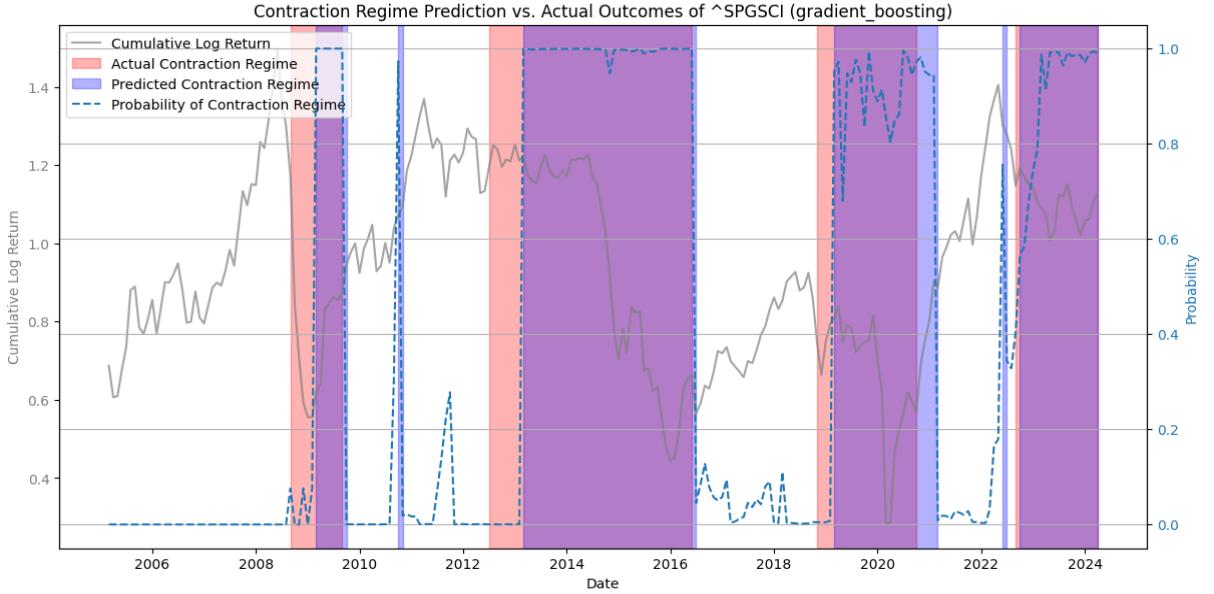


Figure 5.18: Contraction Regimes of GSCI Predicted Through Trend Filtering + Gradient Boosting

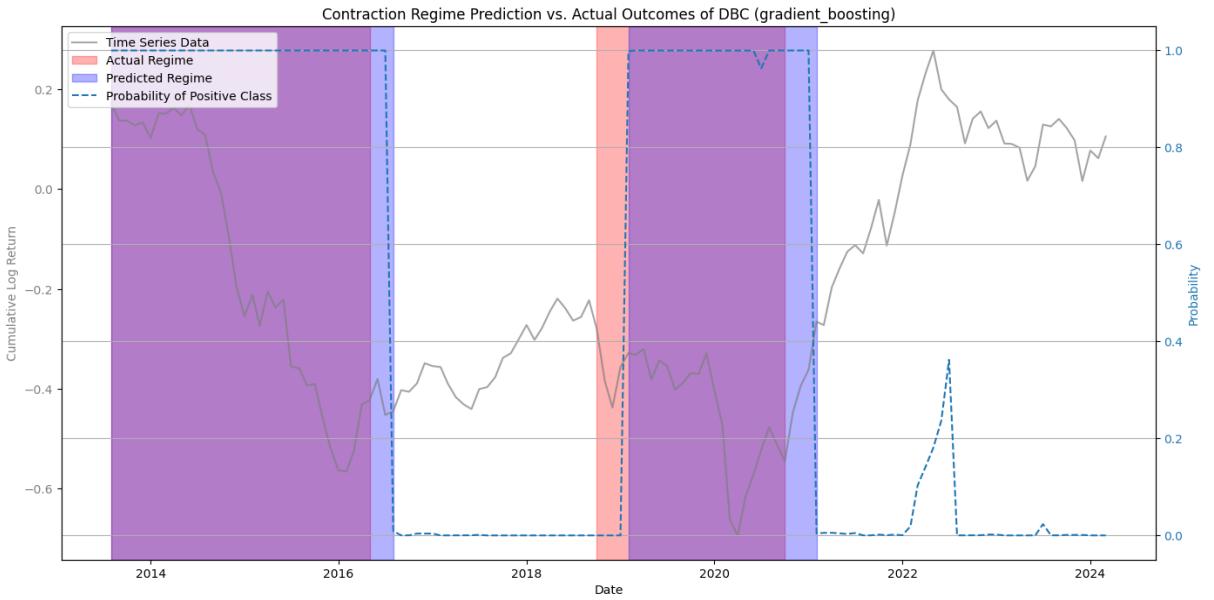


Figure 5.19: Contraction Regimes of DBC Predicted Through Trend Filtering + Gradient Boosting

validation. See Table 5.5 and Table 5.6 for the regime prediction accuracy. The following subsections provide an analysis of the regime prediction outcomes of each of the models when trained with the regime information obtained from jump models.

Model	Test Accuracy (%)
ℓ_1 Logistic Regression	89.13
FFNN	68.26
RNN	54.78
Random Forest	80.43
Gradient Boosting	77.83

Table 5.5: GSCI Regime Prediction Test Accuracy with Regime Information Obtained Through Jump Models

Model	Test Accuracy (%)
ℓ_1 Logistic Regression	84.35
FFNN	60.87
RNN	40.00
Random Forest	70.43
Gradient Boosting	78.26

Table 5.6: DBC Regime Prediction Test Accuracy with Regime Information Obtained Through Jump Models

5.3.1 Logistic Regression

For the GSCI dataset, ℓ_1 logistic regression emerges as the most accurate model, achieving a test accuracy of 89.13%. This high level of accuracy suggests the effectiveness of the ℓ_1 regularization in enhancing the model's generalization capability by preventing overfitting. Although the results seem to be slightly discontiguous, most of the predicted contraction regimes align with the identified regimes, with a small portion of false positives.

Similarly, ℓ_1 logistic regression yields the highest test accuracy for the DBC dataset. Although the regime predictions display discontiguity, the predictions seem to be consistent with the regime identification outcomes with few false negatives.

5.3.2 Neural Networks

The neural network models, FFNN and RNN, show the lowest accuracies of 68.26% and 54.78% for GSCI respectively, indicating their inability to capture the temporal dependencies or overfitting issues due to the excessive number of parameters used for

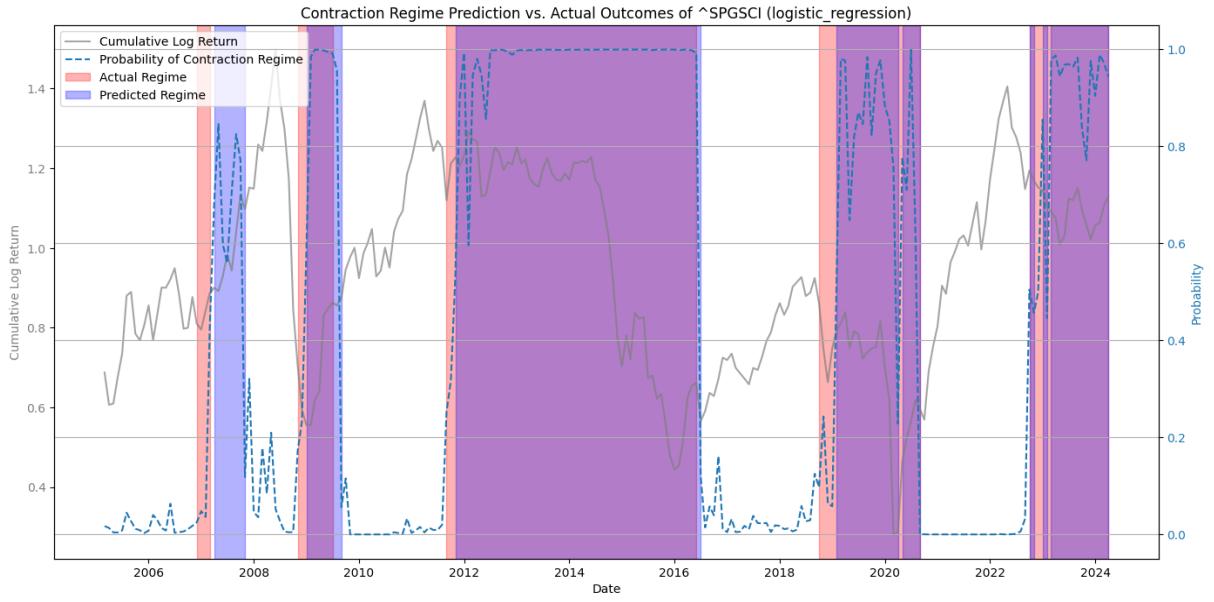


Figure 5.20: Contraction Regimes of GSCI Predicted Through Jump Models + Logistic Regression

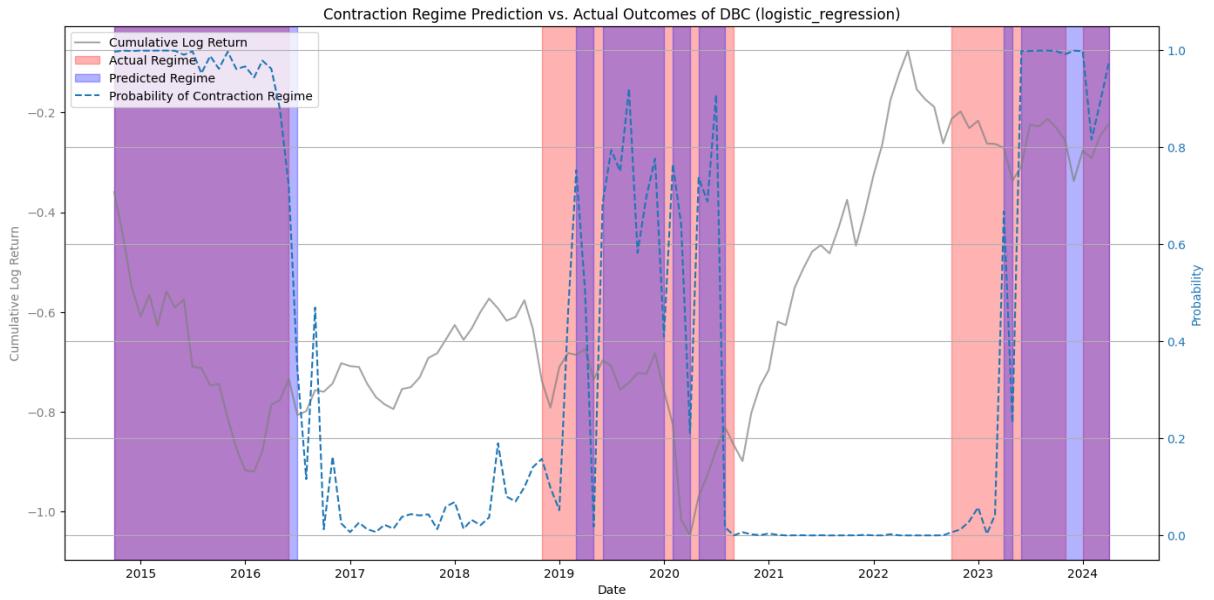


Figure 5.21: Contraction Regimes of DBC Predicted Through Jump Models + Logistic Regression

prediction.

The prediction outcomes on DBC are also disappointing, with test accuracies of 60.87% and 40.00%. In fact, as illustrated in Figure 5.25, RNN's prediction of

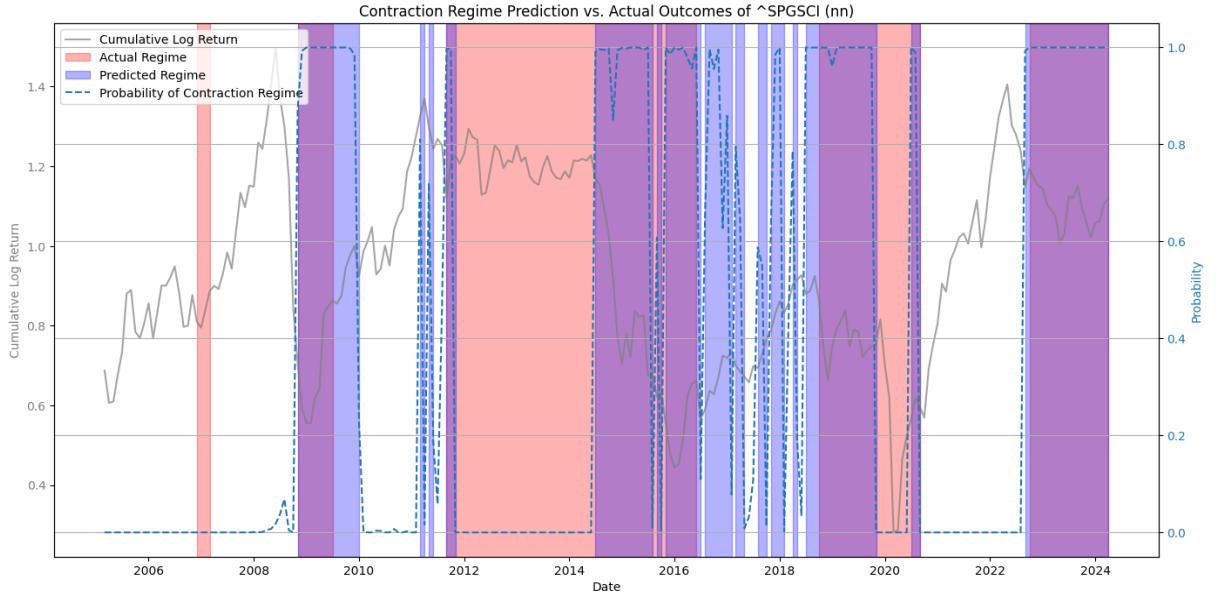


Figure 5.22: Contraction Regimes of GSCI Predicted Through Jump Models + Feed-forward Neural Networks

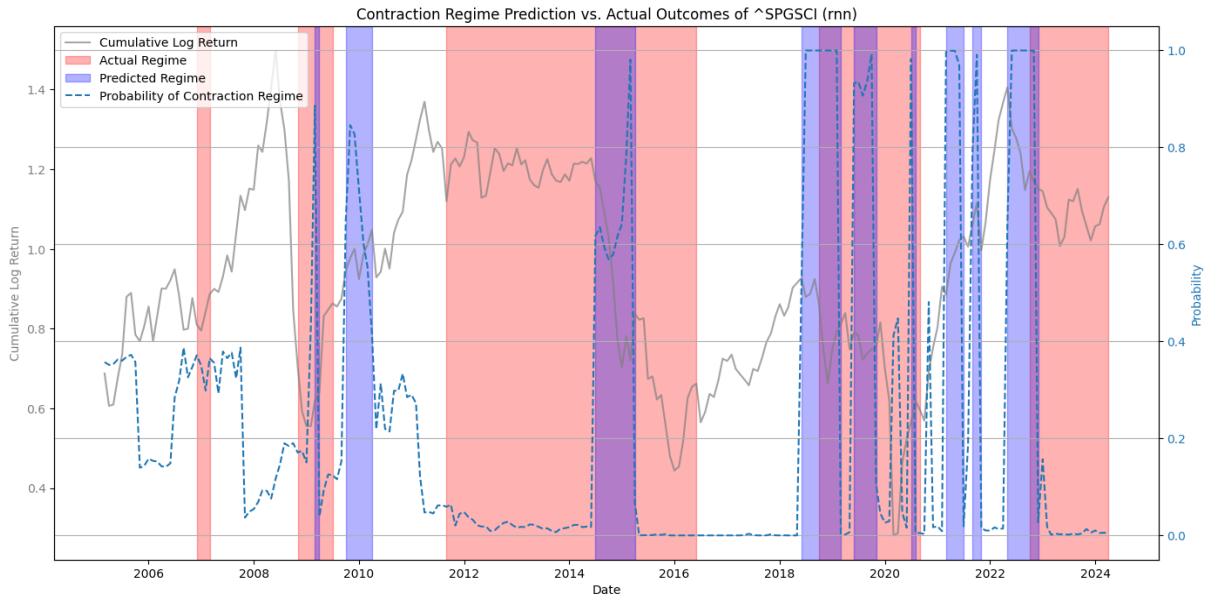


Figure 5.23: Contraction Regimes of GSCI Predicted Through Jump Models + Recurrent Neural Networks

contraction regimes covers almost the entire time period, which is even worse than mere guessing.

Test accuracy is lower on DBC than on GSCI, suggesting that the poorer performance of regime prediction can potentially be attributed to the different nature of

the data or the complexity of the models relative to the smaller data.

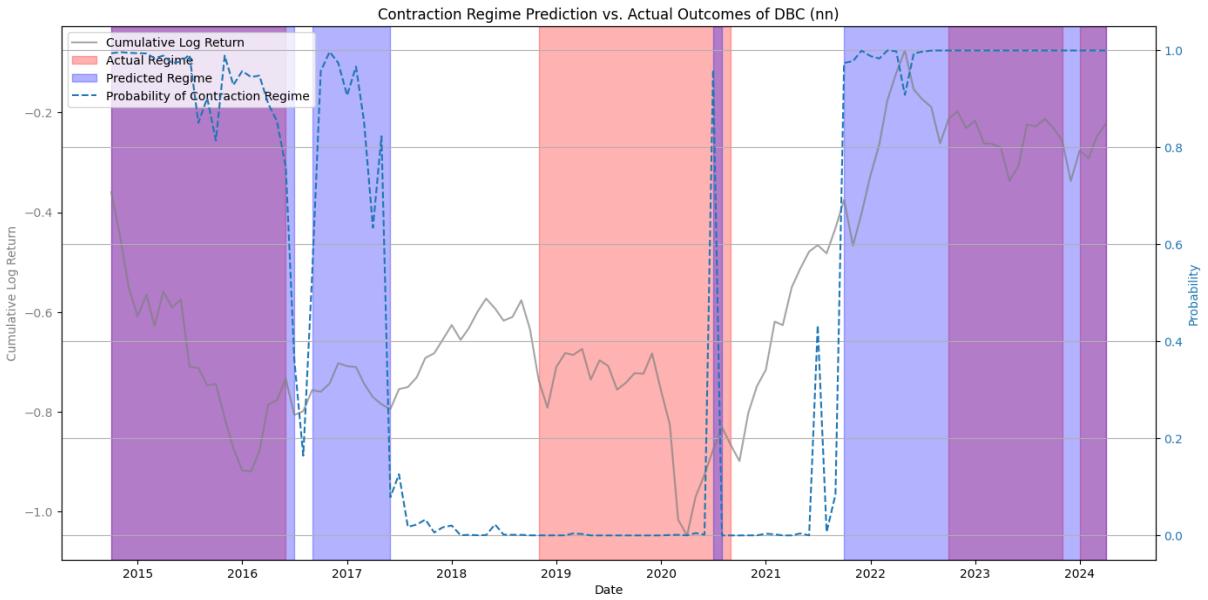


Figure 5.24: Contraction Regimes of DBC Predicted Through Jump Models + Feed-forward Neural Networks

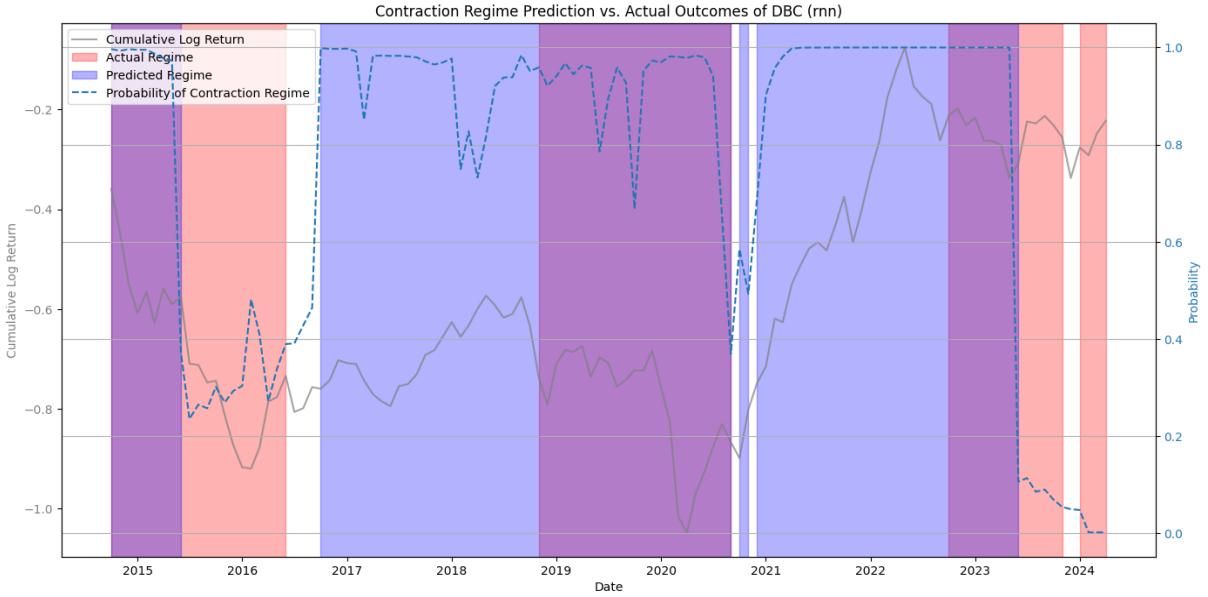


Figure 5.25: Contraction Regimes of DBC Predicted Through Jump Models + Recurrent Neural Networks

5.3.3 Random Forests

The random forest model demonstrates an accuracy of 80.43% for GSCI, ranking the second-highest in accuracy following logistic regression. Despite discrepancies in precisely matching the start and end points of contraction regimes compared to the outcomes of regime identification, the predicted intervals closely parallel the actual ones. This suggests that the observed inaccuracies predominantly stem from minor divergences in data interpretation. However, the model shows a rather low accuracy of 70.43% when applied to the DBC dataset, ranking third out of the five models.

See Figure 5.26 and 5.27 for the graph depiction of the outcomes.

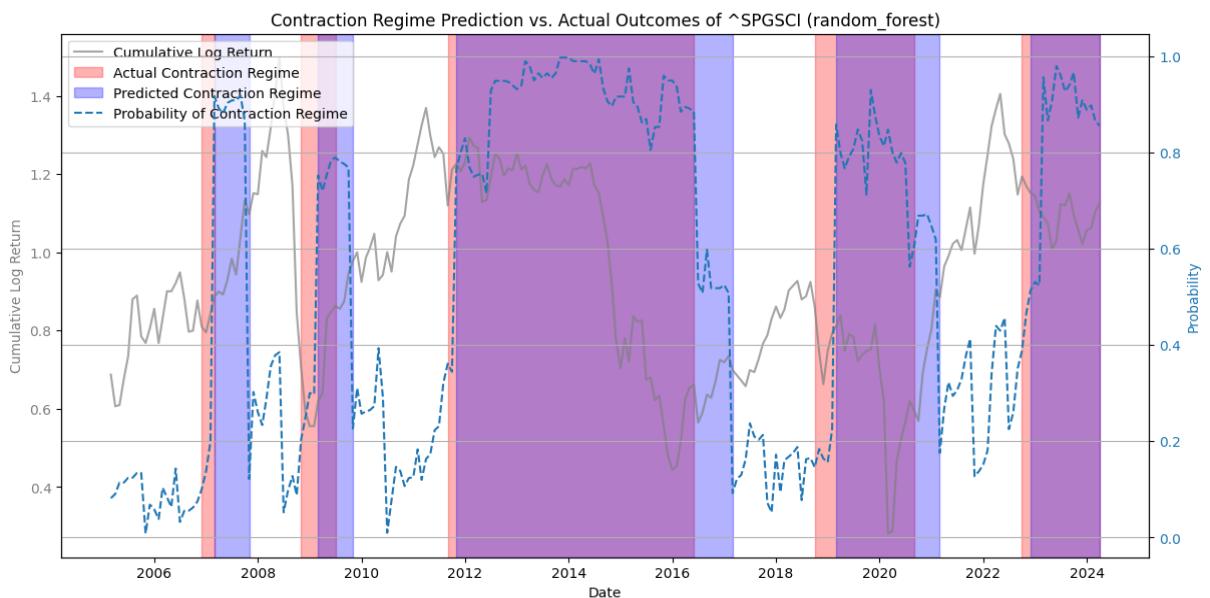


Figure 5.26: Contraction Regimes of GSCI Predicted Through Jump Models + Random Forests

5.3.4 Gradient Boosting

Gradient boosting displays less favorable prediction outcomes when paired with regime identification via jump models compared to when it is paired with regime identification via trend-filtering. As illustrated in Figures 5.28 and 5.29, the predictions are very fragmented, despite a significant fraction of the predicted contraction regimes

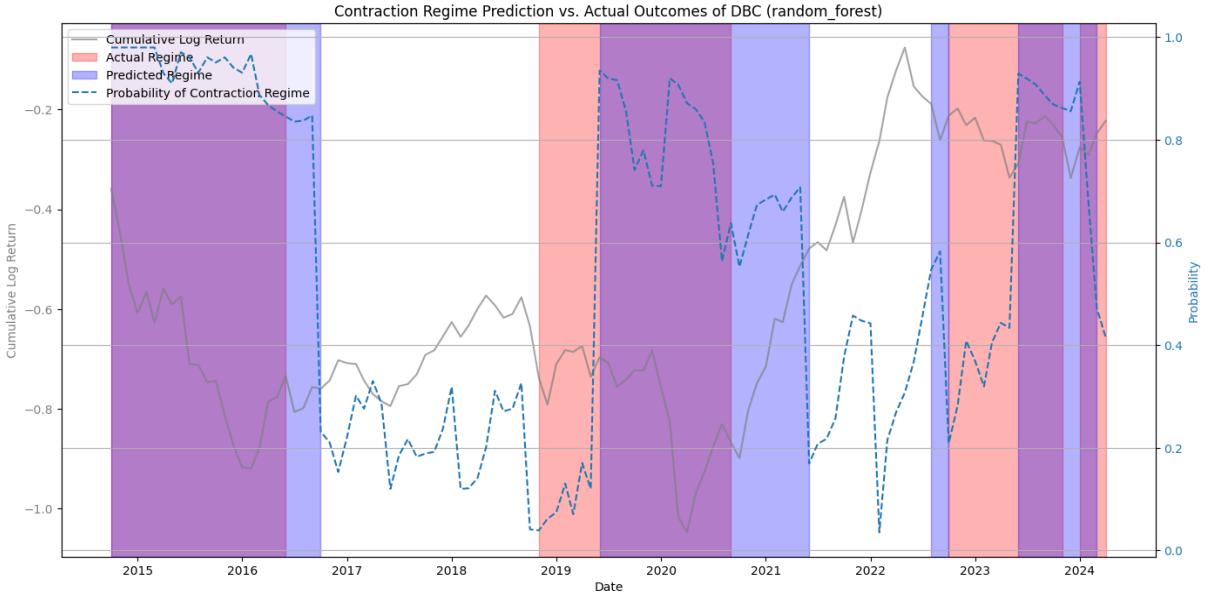


Figure 5.27: Contraction Regimes of DBC Predicted Through Jump Models + Random Forests

aligning with those identified. Furthermore, the model generates numerous false positives, even in periods where cumulative returns evidently show up-trends. This indicates the model's shortcomings in handling the dataset's temporal complexities.

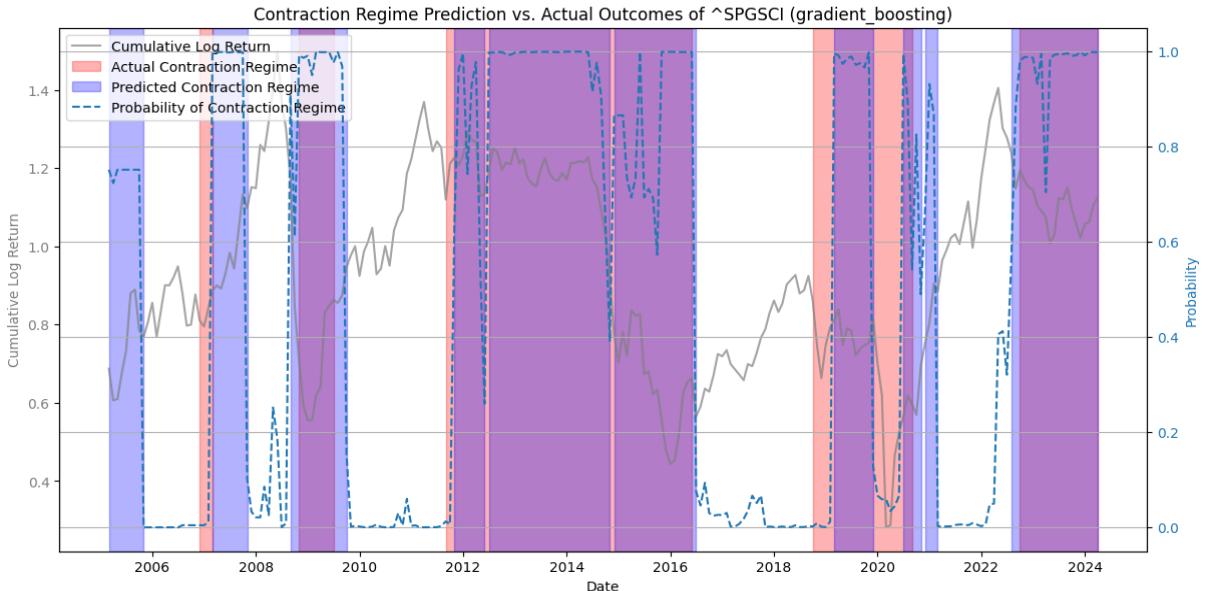


Figure 5.28: Contraction Regimes of GSCI Predicted Through Jump Models + Gradient Boosting

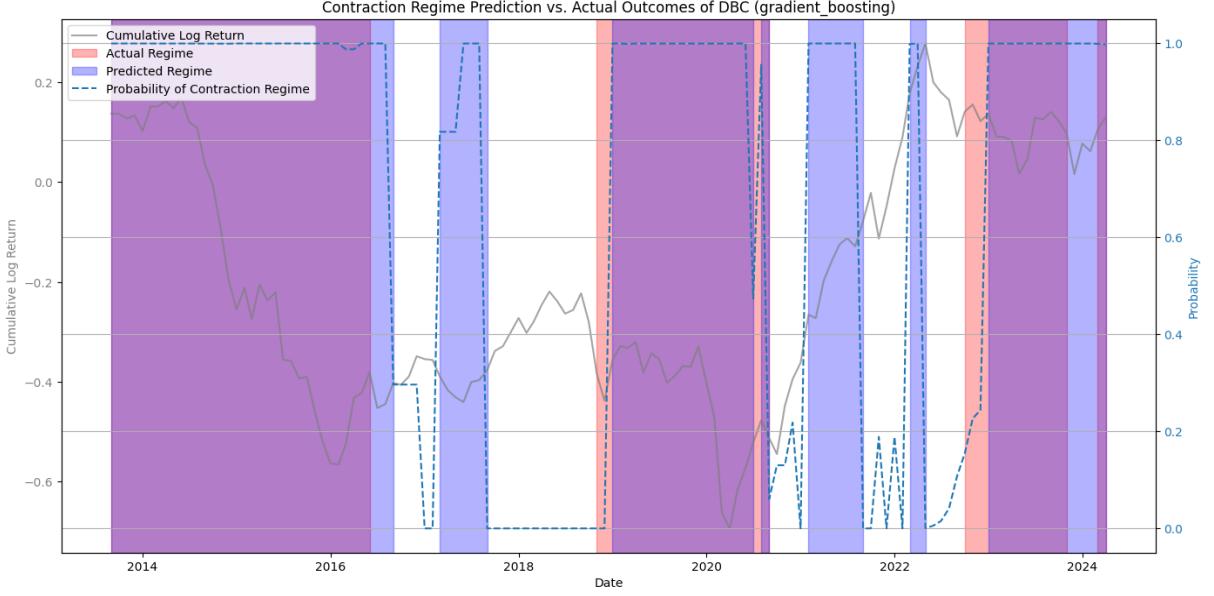


Figure 5.29: Contraction Regimes of DBC Predicted Through Jump Models + Gradient Boosting

5.4 Final Remarks on Regime Prediction

Overall, the test accuracies of regime prediction are higher when the machine-learning models are paired with regime identification via the ℓ_1 trend-filtering algorithm. Also, we notice higher test accuracy in general when the model is tested on the GSCI data, potentially due to more data availability, offering more opportunity to capture the intricacies in the observed data. Although the regime identification outcomes are more contiguous when using jump models, we decide that the regime identification via trend-filtering is more compatible with the regime prediction task.

Meanwhile, there are variations in the performance of each model based on the data it is being tested on. Since ℓ_1 logistic regression shows robust and relatively consistent performance for both the GSCI and DBC dataset, we move on to portfolio construction using logistic regression as our primary model for regime prediction.

5.5 Portfolio Construction

As we decide our primary regime prediction model, we utilize such model to construct portfolios under a regime-switching framework and backtest the portfolios to assess their performances. In this section, we illustrate the outcomes of applying the regime-switching framework to mean reversion, trend-following and risk parity with different composition of assets.

Given that the latest inception date among the assets included in our studied portfolios is December 2014, we choose this date as the start point for our backtesting period across all portfolios and strategies. This approach allows us to directly compare the performance of the portfolios and strategies over an identical timeframe.

5.5.1 Mean Reversion

In this subsection, we present the backtesting outcomes of a mean reversion portfolio strategy applied to various portfolios and compare these outcomes against a 60/40 portfolio that consists of 60 percent stocks and 40 percent bonds only, and an Equal Weight portfolio that allocates the same weight to all of the assets in the portfolio. Specifically we use the SPDR S&P 500 ETF Trust (SPY) and the iShares 20+ Year Treasury Bond ETF (TLT) for the components of the 60/40 portfolio. The performance metrics we to compare the strategies are annualized return (Ret.), volatility (Vol.), Sharpe ratio, Calmar ratio, Sortino ratio, and maximum drawdown (Max. DD).

The analysis is conducted over two rebalancing periods of 6-month and 1-year, with an emphasis on understanding the impact of utilizing a regime-switching approach based on commodity index signals and regime probabilities. Refer to Table 4.1 for the specific components of each portfolio with an assumption of no transaction costs.

First we compare the 60/40 portfolio with Portfolio 1, a portfolio that consists of GSCI, SPY and TLT. See Table 5.7 for the outcomes of the 60/40 portfolio, the Mean Reversion (MR) and the Regime-Switching Mean Reversion (RS-MR) strategies.

	Ret. (%)	Vol. (%)	Sharpe	Calmar	Sortino	Max DD. (%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	7.66	10.36	0.74	0.22	0.90	34.46
6-month rebalancing period						
MR	5.53	14.69	0.38	0.21	0.49	26.67
RS-MR	8.03	9.86	0.81	0.41	1.09	19.35
1-year rebalancing period						
MR	5.53	14.69	0.38	0.21	0.49	26.67
RS-MR	7.77	9.63	0.81	0.41	1.13	18.74

Table 5.7: Portfolio 1 Mean Reversion Performance - GSCI

When using the GSCI as the primary commodity index, the RS-MR strategy employed with both rebalancing periods outperforms all other portfolios in terms of annualized return, Sharpe ratio, Calmar ratio, Sortino ratio, and maximum drawdown. These outcomes affirm the RS-MR strategy's adeptness at navigating volatile markets while safeguarding against significant losses. The traditional 60/40 portfolio also shows solid performance with a significant annualized return of 7.72%, and Sharpe and Sortino ratios comparable to those of the RS-MR strategy. However, the maximum drawdown is 26.22%, significantly higher than the 18.74% (6-month rebalance) and 19.35% (1-year rebalance) maximum drawdown observed in the RS-MR portfolio.

In contrast, the MR strategy underperforms across every performance metric in comparison, suggesting that simple mean reversion approach is a less attractive option to investors than the traditional 60/40 portfolio. Meanwhile, the Equal Weight portfolio demonstrates outcomes that are on par with those of 60/40 and RS-MR in nearly every performance metric, even achieving a Sharpe that surpasses the 60/40. However, it also yields the highest maximum drawdown of 34.46%, indicating the increased drawdown risk associated with this portfolio selection.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.90	9.77	0.60	0.18	0.74	32.45
6-month rebalancing period						
MR	5.01	11.88	0.42	0.26	0.63	19.50
RS-MR	8.12	9.29	0.87	0.51	1.25	15.86
1-year rebalancing period						
MR	5.01	11.88	0.42	0.26	0.63	19.50
RS-MR	7.75	9.28	0.84	0.50	1.24	15.48

Table 5.8: Portfolio 1 Mean Reversion Performance - DBC

As we replace the GSCI with the DBC index and conduct backtesting, we likewise discover that the RS-MR strategy outperforms all the other portfolios in every metric for both rebalancing periods, while the MR portfolio shows the most unimpressive performance. Another noteworthy observation is that while all the other portfolios exhibit significantly high maximum drawdowns of approximately 20 percent or higher, the RS-MR portfolio shows the lowest drawdowns of approximately 15–16 percent for both rebalancing periods, demonstrating its effectiveness in limiting losses during market downturns.

Next, we add the S&P Global Broad Market Index (SBBMGLU) to Portfolio 1 and assess its performance against that of the 60/40 benchmark portfolio and the Equal Weight portfolio during the same period. See Table 5.9 for the performance statistics with GSCI as the primary commodity index, and Table 5.10 for the performance statistics based on DBC.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.70	11.47	0.69	0.29	1.00	26.22
Equal Weight	6.63	9.25	0.72	0.45	1.12	14.87
6-month rebalancing period						
MR	6.58	8.97	0.73	0.41	1.09	16.16
RS-MR	10.15	10.12	1.00	0.50	1.49	20.16
1-year rebalancing period						
MR	6.40	10.27	0.62	0.32	0.87	19.77
RS-MR	5.68	9.82	0.58	0.27	0.71	21.36

Table 5.9: Portfolio 2 Mean Reversion Performance - GSCI

The RS-MR strategy outperforms the MR strategy in every performance metric for both rebalancing periods with a slightly higher volatility level for the 6-month rebalancing period. The maximum drawdown values are also slightly higher than those of MR, but the significantly higher Sharpe, Calmar, and Sortino ratios outweigh such slight differences in risk levels, indicating a more efficient risk-adjusted return.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.32	8.48	0.74	0.50	1.30	12.57
6-month rebalancing period						
MR	6.35	8.51	0.75	0.42	1.18	15.01
RS-MR	9.48	9.39	1.01	0.59	1.53	16.06
1-year rebalancing period						
MR	6.11	9.68	0.63	0.33	0.94	18.28
RS-MR	6.21	9.68	0.64	0.37	0.85	16.81

Table 5.10: Portfolio 2 Mean Reversion Performance - DBC

For the DBC dataset, the RS-MR strategy again yields superior performance with a Sharpe Ratio of 1.01 and 0.64 for the respective rebalancing periods of 6 months and 1 year, outperforming the MR strategy's Sharpe ratios of 0.75 and 0.63. The Calmar ratios and Sortino ratios are also mostly higher, further validating the RS-MR strategy's superior risk-adjusted performance.

When compared to the traditional 60/40 and Equal Weight benchmark portfolios, the RS-MR strategy demonstrates solid performance across both datasets and rebalancing periods. The RS-MR seems to be very effective especially in mitigating large losses compared to the 60/40 portfolio, considering the big difference in the maximum drawdown value. RS-MR outshines the other portfolios especially when coupled with a 6-month rebalancing period, yielding the highest performance metric.

Then, we move on to testing the performance of Portfolio 3, a portfolio that consists of the primary commodity index (GSCI or DBC), SPY, TLT, along with iShares 7–10 Year Treasury Bond ETF (IEF), iShares 1–3 Year Treasury Bond ETF (SHY), and Invesco S&P SmallCap Value with Momentum ETF (XSVM). The results

are depicted in Table 5.11 and Table 5.12.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.41	8.94	0.61	0.35	0.78	15.43
6-month rebalancing period						
MR	6.82	8.67	0.78	0.37	0.98	18.01
RS-MR	7.94	7.90	1.00	1.36	1.35	5.81
1-year rebalancing period						
MR	4.95	6.96	0.71	0.34	0.83	14.14
RS-MR	7.68	8.24	0.93	1.31	1.12	5.81

Table 5.11: Portfolio 3 Mean Reversion Performance - GSCI

Compared to the MR strategy, the RS-MR strategy achieves a superior annualized return of 7.94% with a lower volatility of 7.90% for the 6-month rebalancing period and a significantly greater annualized return of 7.68% with slightly higher volatility of 8.24% for the 1-year rebalancing period, ultimately exhibiting higher Sharpe, Calmar and Sortino ratios. Such outcomes underscore the RS-MR strategy's effectiveness in navigating market volatilities and drawdowns. The RS-MR likewise outperforms the benchmark portfolios in every performance metric. In this scenario, the RS-MR demonstrates its exceptional ability to manage drawdown risks using the regime information incorporated into the strategy.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.12	8.24	0.74	0.25	0.87	24.50
6-month rebalancing period						
MR	4.96	9.59	0.52	0.32	0.72	15.36
RS-MR	7.44	7.71	0.96	0.41	0.94	13.96
1-year rebalancing period						
MR	5.81	7.54	0.77	0.45	1.00	12.86
RS-MR	5.81	7.55	0.77	0.32	0.72	18.21

Table 5.12: Portfolio 3 Mean Reversion Performance - DBC

On the DBC dataset, the RS-MR strategy again demonstrates higher Sharpe Ratios—0.96 and 0.77—than those of the MR strategy—0.52 and 0.77. Its Calmar and

Sortino ratios are also higher with a 6-month rebalancing period, indicating its robustness in managing downside risk and capitalizing on mean reversion opportunities. However, the Calmar and Sortino values of a 1-year rebalancing RS-MR are lower than those of the 1-year rebalancing MR, implying that the longer rebalancing interval may dilute the advantage of the RS-MR, which is its prompt response to market regime changes.

Across both datasets and rebalancing intervals, the RS-MR strategy consistently outperforms both the 60/40 and Equal Weight benchmark portfolios in all considered risk-adjusted return metrics. One notable fact is that the maximum drawdown is significantly lower than the 60/40 portfolio, which highlights the RS-MR strategy's superiority in managing drawdowns effectively and reducing the impact of downturns on the portfolio's performance.

Lastly, we integrate the iShares 1–5 Year Investment Grade Corporate Bond ETF (IGSB) and the BMI index, showcasing a portfolio that spans small-cap to large-cap equities, domestic to global equities, and a mix of corporate and government bonds, further with the inclusion of the commodity index.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.35	7.28	0.74	0.44	0.97	12.21
6-month rebalancing period						
MR	5.01	7.15	0.70	0.36	0.91	13.92
RS-MR	6.52	7.21	0.90	0.47	1.20	13.92
1-year rebalancing period						
MR	5.86	9.04	0.65	0.35	0.83	16.51
RS-MR	5.86	9.04	0.65	0.36	0.69	16.51

Table 5.13: Portfolio 4 Mean Reversion Performance - GSCI

For the 6-month rebalancing period, the RS-MR strategy exhibits superior performance compared to the MR strategy, yielding a higher annualized return of 6.52% against 5.01% for GSCI and 6.32% against 5.09% for DBC with a slightly higher or equal volatility of 7.21% (RS-MR, GSCI) and 6.90% (RS-MR, DBC) against 7.15%

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.21	6.87	0.76	0.45	1.08	11.58
6-month rebalancing period						
MR	5.09	6.91	0.74	0.38	0.99	13.57
RS-MR	6.32	6.90	0.92	0.50	1.20	12.60
1-year rebalancing period						
MR	5.08	7.81	0.65	0.40	0.76	12.59
RS-MR	5.72	8.68	0.66	0.36	0.88	15.97

Table 5.14: Portfolio 4 Mean Reversion Performance - DBC

(MR, GSCI) and 6.91% (MR, DBC). This strategy also achieves higher Sharpe, Calmar and Sortino ratios, showing greater resilience against drawdowns.

With the 1-year rebalancing period, both strategies display approximately the same annualized returns around 5.86% with GSCI as the primary commodity index. For DBC, the Calmar ratio is actually higher for MR due to the lower drawdown value of 12.59% against 15.97%, but the RS-MR strategy demonstrates a higher Sharpe and Sortino, indicating that the RS-MR strategy achieves better risk-adjusted returns and more effectively manages volatility and downside risk, despite experiencing a slightly higher maximum drawdown. Compared to the traditional 60/40 and Equal Weight benchmarks, the RS-MR strategy shows higher Sharpe, Calmar and Sortino ratios with the 6-month rebalancing period, while the metrics generally underperform the 60/40 portfolio with a 1-year rebalance.

From the outcomes, we see a general trend of better performance of the RS-MR strategy with shorter rebalancing period, which is especially accentuated for Portfolio 4 containing the most diverse set of assets. Such superiority could stem from many different factors. One potential reason could be the greater diversity of assets being able to quickly capture the benefits of regime-switching mean reversion over shorter intervals, due to mean reversion phenomenon being more evident from short-term market inefficiencies before the market is corrected long-term. Such mean reversion can be exploited by a well-diversified portfolio. In fact, the market's mean reverting

tendency might be diluted over time, especially with a broader asset base, so there will be less impact on the portfolio performance even if the investor correctly identifies mean reversion opportunities. This could potentially explain the poorer performance compared to the 60/40 strategy, which benefits from long-term trends.

The analysis highlights the advantages of integrating the regime-switching framework into mean reversion strategies. The RS-MR strategy not only achieves higher returns across both the GSCI and DBC datasets but also demonstrates higher risk-adjusted performance compared to its MR counterpart. Also, the RS-MR strategy is more effective than the benchmark portfolios, with higher performance in Sharpe, Calmar and Sortino ratios, suggesting that it is more adept at capitalizing on the nature of financial markets and mitigating losses during downturns. These findings ultimately underscore the potential of dynamic, regime-dependent strategies in enhancing portfolio performance and resilience.

5.5.2 Trend-Following

Now we examine the effectiveness of applying the regime-switching framework to trend-following strategies. Specifically, we backtest the performance of the Trend-Following (TF) and Regime-Switching Trend-Following (RS-TF) strategies as well as the benchmark portfolios, 60/40 and Equal Weight, across two commodity indices: GSCI and DBC. This time, we do not divide the outcomes of TS and RS-TF based on the rebalancing period, since rebalancing occurs monthly based on the trend-following signal. Our goal is to assess and compare the strategies' performances under varying market conditions using the same set of performance metrics.

The RS-TF strategy outperforms all other strategies with an impressive annualized return of 9.62% and a Sharpe Ratio of 0.93, significantly higher than all the other portfolios. This superior performance, coupled with the highest Sortino and Calmar ratios, demonstrates the RS-TF strategy's effectiveness in exploiting the market

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	7.66	10.36	0.74	0.22	0.90	34.46
TF	6.85	12.30	0.56	0.34	0.76	20.21
RS-TF	9.62	10.36	0.93	0.49	1.21	19.70

Table 5.15: Portfolio 1 Trend-Following Performance - GSCI

trends for profit while managing downside risks efficiently. The maximum drawdown of RS-TF is lowest among all the portfolios, which exhibits high maximum drawdown values ranging from approximately 20 to 35 percent, making the RS-TF strategy appealing for risk-averse investors seeking to minimize losses during periods of market turmoil.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.90	9.77	0.60	0.18	0.74	32.45
TF	6.51	10.75	0.61	0.41	0.91	15.80
RS-TF	8.93	9.39	0.95	0.55	1.19	16.14

Table 5.16: Portfolio 1 Trend-Following Performance - DBC

Similar to the GSCI dataset, the RS-TF strategy in the DBC dataset presents a significantly high annualized return of 8.93% and a Sharpe Ratio of 0.95, surpassing the equivalent metrics of the traditional TF strategy and benchmark portfolios. The strategy also interestingly exhibits a maximum drawdown higher than that of TF, suggesting that the RS-TF might allocate weights to each asset of a portfolio in a way that increases vulnerability to specific market events such as trend reversals or market shocks.

Across both datasets, the RS-TF strategy consistently yields higher risk-adjusted performance metrics compared to the traditional TF strategy and benchmark portfolios. The strategy seems to be successful in navigating the volatile nature of commodity markets by incorporating the regime-switching framework.

Now we move on to analyzing the TF and RS-TF strategy's effectiveness on Port-

folio 2, an addition of S&P Global Broad Market Index (SBBMGLU) to Portfolio 1.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.70	11.47	0.69	0.29	1.00	26.22
Equal Weight	6.63	9.25	0.72	0.45	1.12	14.87
TF	7.41	12.24	0.61	0.37	0.82	20.21
RS-TF	9.16	10.18	0.90	0.46	1.17	19.70

Table 5.17: Portfolio 2 Trend-Following Performance - GSCI

For the GSCI, the RS-TF strategy remains consistently effective, yielding the highest Sharpe, Calmar and Sortino ratios. All the performance metrics are remarkably higher than the performance metrics of the TF strategy, indicating that the regime-switching approach significantly improves the performance of the traditional trend-following strategy.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.32	8.48	0.74	0.50	1.30	12.57
TF	6.93	10.73	0.65	0.44	0.96	15.80
RS-TF	8.54	9.21	0.93	0.53	1.19	16.14

Table 5.18: Portfolio 2 Trend-Following Performance - DBC

The Sharpe ratio of RS-TF is highest for the DBC as well, although the Equal Weight portfolio shows better performance in terms of the Sortino ratio. RS-TF again outperforms TF in every performance metric, illustrating the importance of the regime-switching approach in the RS-TF strategy.

When compared to Portfolio 1, Portfolio 2 exhibits lower returns and volatilities, potentially due to the introduction of diversification benefits caused by the addition of the S&P Global BMI Index. Diversification can moderate portfolio volatility, though it may also slightly dilute overall profit generation capability.

Now we move on to assessing the performance of RS-TF and TF when applied to Portfolio 3, and compare them against the benchmarks.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.41	8.94	0.61	0.35	0.78	15.43
TF	6.04	10.00	0.60	0.37	0.78	16.47
RS-TF	7.51	8.72	0.86	0.55	1.13	13.63

Table 5.19: Portfolio 3 Trend Following Performance - GSCI

The RS-TF strategy significantly outperforms TF and the benchmark portfolios with GSCI as the primary commodity index, boasting the highest Sharpe ratio of 0.86, Calmar ratio of 0.55 and Sortino ratio of 1.13. Such outcomes indicate a robust risk-adjusted performance. The RS-TF strategy also reduces the maximum drawdown to 13.63%, the lowest among all four portfolios, highlighting its effectiveness in safeguarding against large-scale losses during extreme market downturns.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.12	8.24	0.74	0.25	0.87	24.50
TF	5.84	9.11	0.64	0.41	0.89	14.14
RS-TF	7.14	8.17	0.87	0.57	1.14	12.46

Table 5.20: Portfolio 3 Trend Following Performance - DBC

Similar to the GSCI dataset, the RS-TF strategy in the DBC dataset outperforms the other portfolios with the highest Sharpe, Calmar and Sortino ratios as well as a maximum drawdown considerably lower than those of other portfolios, indicating that the RS-TF strategy's strong performance is consistent across different commodity index datasets.

Lastly, we conduct backtesting using Portfolio 4, a portfolio that consists of a comprehensive set of asset classes with the most diversification. See Table 5.21 and Table 5.22 for the performance metrics computed in the backtesting process.

In the GSCI dataset, the RS-TF strategy displays a lower annualized return than the 60/40 portfolio but also lower volatility, leading to a higher Sharpe ratio. RS-TF's performance metrics excel its TF counterpart as well, showcasing its strength in man-

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.35	7.28	0.74	0.44	0.97	12.21
TF	5.81	9.02	0.64	0.39	0.81	14.76
RS-TF	6.88	7.81	0.88	0.56	1.16	12.29

Table 5.21: Portfolio 4 Trend Following Performance - GSCI

aging downside risks while yielding high returns. The RS-TF strategy is overall most adept at navigating market volatility while also capitalizing on market opportunities to generate profit.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.21	6.87	0.76	0.45	1.08	11.58
TF	5.56	8.26	0.67	0.43	0.91	12.93
RS-TF	6.55	7.32	0.89	0.58	1.18	11.19

Table 5.22: Portfolio 4 Trend Following Performance - DBC

In the DBC dataset, the RS-TF strategy again stands out with the highest Sharpe, Calmar and Sortino ratios, along with second-highest annualized return. It clearly surpasses the Equal Weight and TF portfolios in performance, delivering higher returns with lower maximum drawdowns, while maintaining comparable or even lower volatility. Also, although the return (6.55%) might be slightly lower than that of 60/40 (7.72%), RS-TF shows substantially lower volatility, underscoring the RS-TF strategy's efficiency in generating higher returns per unit of risk taken.

Across Portfolios 1, 2, 3, and 4, a consistent pattern emerges where the RS-TF outperforms the TF strategy as well as the benchmark portfolios. Although Portfolio 1 performs better than Portfolio 2, a portfolio with more diversification, as we add more diversification factors into the portfolio, we observe a trend of the portfolio's performance improving with increased diversification. This indicates the strategy's capability to leverage diversification while maintaining strong performance metrics.

Such observations can be attributed to the RS-TF strategy's capability to dynam-

ically adapt to market conditions to more effectively pursue profit and manage risks than static strategies. Considering the volatile nature of commodities market, the RS-TF strategy may be even crucial to navigate the market volatility. Meanwhile, the differences in performance across portfolios may be reflective of the different degrees of diversification and market sensitivity based on the composition of the portfolio. In the context of trend-following, we notice the pattern of increased diversification with the addition of market indices leading to lower maximum drawdowns but also reduced potential to maximize profits due to increased exposure to different markets of bonds, equities and commodities.

We conclude this subsection by stating that the analysis demonstrates the superior performance of the Regime-Switching Trend-Following (RS-TF) strategy across the GSCI and DBC datasets, and across different portfolios with different asset components, compared to traditional trend-following approaches and benchmark portfolios. This strategy's promising results advocate for its potential to enhance the performance of a diversified investment portfolio, offering a novel opportunity to improve returns and manage risks in volatile commodity markets.

5.5.3 Risk Parity

In this subsection, we present the results and analysis of risk parity portfolios, the traditional Risk Parity (RP) and the Regime-Switching Risk Parity (RS-RP), each incorporating a different commodity index, the GSCI and DBC. Again, we compare the performance against those of benchmark portfolios. See Table 5.23 and 5.24 for the performance of Portfolio 1.

The RS-RP strategy with GSCI yields an annual return of 4.56%, a return higher than the 3.62% yielded by the RP strategy, and also shows higher Sharpe, Calmar and Sortino ratios. RP and RS-RP strategies both underperform the 60/40 and Equal Weight portfolios in terms of all peformance metrics, although the strategies exhibit

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	7.66	10.36	0.74	0.22	0.90	34.46
RP	3.62	9.73	0.37	0.13	0.56	27.91
RS-RP	4.56	9.81	0.46	0.17	0.68	26.67

Table 5.23: Portfolio 1 Risk Parity Performance - GSCI

maximum drawdown lower than that of the Equal Weight portfolio.

The Equal Weight portfolio boasts the highest ratio of 0.74, while the 60/40 boasts the highest Calmar and Sortino value of 0.29 and 1.00 respectively. This outcome suggests that although RP and RS-RP are intended to spread risk evenly, the benchmark portfolios may actually perform better since RP and RS-RP inevitably leads to a heavier concentration in certain asset classes. Such high concentration could cause huge drawdowns in the portfolio during occasions in which the heavily concentrated assets experience market downturns.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.90	9.77	0.60	0.18	0.74	32.45
RP	3.41	9.14	0.37	0.15	0.56	22.99
RS-RP	4.44	9.34	0.48	0.20	0.72	22.02

Table 5.24: Portfolio 1 Risk Parity Performance - DBC

When we include DBC in Portfolio 1, RS-RP again outperforms the RP in all performance metrics. The RS-RP strategy with DBC shows a Calmar ratio higher than that of the Equal Weight portfolio, with a comparable value of Sortino. Also, the maximum drawdown of RS-RP is significantly lower than that of the Equal Weight portfolio. Meanwhile, the RS-RP underperforms the 60/40 portfolio in the three ratios, although the maximum drawdown is lower. The maximum drawdowns of the 60/40 and Equal Weight portfolio mark high rates above 25 percent, but we notice that the RP and RS-RP portfolios are less impacted by such market events, an observation that advocates for the advantage of investing in commodity futures

under a regime-switching approach for diversification to mitigate losses from market crashes.

Overall, the analysis of Portfolio 1 indicates that the RS-RP strategy performs better than the RP strategy. Also, we notice that the RP and RS-RP strategies do not necessarily outperform the 60/40 or Equal Weight portfolios in terms of annualized returns or the Sharpe, Calmar and Sortino ratios. In fact, under certain market conditions, it almost seems more promising to employ the traditional portfolios due to higher opportunities to capture the upside during market rallies and gain high risk-adjusted returns. However, these two strategies, especially the RS-RP strategy, show robust results in drawdown management and bolster the diversification benefits of commodities, making them attractive options during volatile market conditions by reducing the dependence on a particular asset class associated with higher risk.

Now we analyze Portfolio 2, adding the S&P Global BMI Index to Portfolio 1. Refer to Table 5.25 and 5.26 for specific outcomes.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.63	9.25	0.72	0.45	1.12	14.87
RP	6.27	11.36	0.55	0.43	1.76	14.52
RS-RP	5.29	6.63	0.80	0.48	1.54	10.92

Table 5.25: Portfolio 2 Risk Parity Performance - GSCI

With GSCI, RS-RP exhibits the lowest return, but also the highest Sharpe and Calmar ratios among the four strategies due to the strategy's lowest volatility and maximum drawdown. Its Sortino ratio is lower than RP, indicating that RS-RP yields better risk-adjusted performance but may be less effective in managing downside risk. RP yields a higher volatility but also higher annualized returns, showing that the strategy is able to translate its high volatility into higher return.

The 60/40 portfolio shows the highest return in exchange for the highest maximum drawdown and volatility, implying that although the portfolio may yield high returns,

it is also vulnerable to market volatility. The Equal Weight portfolio is outperformed by the RS-RP in all three ratios, but outperforms RP in terms of the Sharpe and Calmar ratios.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.32	8.48	0.74	0.50	1.30	12.57
RP	5.76	10.91	0.53	0.47	1.69	12.23
RS-RP	5.10	6.68	0.76	0.46	1.58	11.07

Table 5.26: Portfolio 2 Risk Parity Performance - DBC

When we replace GSCI with DBC, the 60/40 portfolio again leads with the highest annualized return, followed by the 6.32% return of the Equal Weight portfolio. However, the RS-RP stands out with its highest Sharpe ratio due to its lowest volatility level, which is significantly lower than the volatilities of other portfolios. RP shows robust performance in terms of Sortino ratio, but it lags with a Sharpe ratio of 0.53.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	6.12	8.24	0.74	0.25	0.87	24.50
RP	1.39	1.78	0.78	0.27	0.98	5.20
RS-RP	1.22	1.83	0.67	0.23	0.90	5.32

Table 5.27: Portfolio 3 Risk Parity Performance - GSCI

Now we move on to examining the backtesting outcomes of Portfolio 3. Both RP and RS-RP offer minimal annualized returns of 1.39% and 1.22%, significantly lower than the returns of the benchmark portfolios, but also a very low volatility of approximately 1.8%. The RP interestingly shows a high Sortino ratio following that of the 60/40, an outcome that underscores its effectiveness in mitigating downside risk. Conversely, the RS-RP strategy, with its slightly lower returns and higher maximum drawdown compared to RP, highlights the challenge of seeking returns while controlling for drawdown risk.

The 60/40 portfolio shows the highest annualized return as well as the highest

Sharpe, Calmar and Sortino. Yet, it is noteworthy that the maximum drawdown is significantly higher than that of RP and RS-RP, indicating a higher risk of substantial losses during market downturns.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD. (%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.60	7.94	0.70	0.24	0.84	23.30
RP	1.39	1.81	0.76	0.27	0.99	5.17
RS-RP	1.24	1.83	0.63	0.22	0.91	5.49

Table 5.28: Portfolio 3 Risk Parity Performance - DBC

As we analyze the results after replacing GSCI with DBC, we notice that the RP strategy yields not only the highest return but also the highest Sharpe ratio. Both RP and RS-RP mirror their effectiveness in drawdown management, with their maximum drawdowns being significantly lower than those of 60/40 and Equal Weight. The RP exceeds RS-RP in every key performance metric (Sharpe, Calmar, and Sortino ratios), reaffirming the previous findings from Portfolio 3 with GSCI.

Lastly, we move on to the analysis of Portfolio 4, a portfolio including a comprehensive set of asset classes and thus the greatest degree of diversification.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD. (%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.35	7.28	0.74	0.44	0.97	12.21
RP	2.14	2.53	0.76	0.33	1.54	6.43
RS-RP	1.95	2.53	0.77	0.25	1.22	7.69

Table 5.29: Portfolio 4 Risk Parity Performance - GSCI

Here, we notice that the RS-RP has a slight edge over RP in Sharpe despite lower returns, which could be attributed to lower volatility. However, it is outperformed by RP in terms of every other performance metric. Both RP and RS-RP offer very low volatility and maximum drawdown compared to the benchmark portfolios, underscoring its conservative risk profile. The RP strategy stands out with the highest Sortino ratio on top of its lowest maximum drawdown, showing its effectiveness in

managing downside risk.

The 60/40 strategy, while delivering the highest returns, also yields the highest volatility, resulting in the lowest Sharpe ratio. Also, it bears the highest maximum drawdown. The Equal Weight strategy boasts the highest Calmar ratio of 0.44.

	Ret.(%)	Vol.(%)	Sharpe	Calmar	Sortino	Max DD.(%)
60/40	7.72	11.41	0.68	0.29	1.00	26.22
Equal Weight	5.21	6.87	0.76	0.45	1.08	11.58
RP	1.63	2.03	0.80	0.26	1.17	6.40
RS-RP	2.01	2.51	0.80	0.27	1.28	7.40

Table 5.30: Portfolio 4 Risk Parity Performance - DBC

Now we shift focus to the DBC dataset and assess Portfolio 4's performance. This time, we notice that the RS-RP yields higher return and Calmar ratio than RP. RS-RP in fact shows the highest Sharpe, Calmar and Sortino ratios among the four due to its low maximum drawdown and volatility, implying that it effectively manages the portfolio's risk during market downturns. The performance metrics of RP are similar to those of RS-RP, indicating that the addition of regime-switching to the risk parity strategy offers marginal improvements in this context. RP and RS-RP excel the benchmarks in their Sharpe and Sortino ratios, although they yield lower Calmar ratios.

Overall, we notice that with high diversification resulting from the inclusion of many asset classes, the RP and RS-RP strategies yield returns that are modest compared to the traditional portfolios since they prioritize risk control. However, the two strategies deliver lower volatility, leading to a moderately high risk-adjusted return. Also, these strategies excel in minimizing drawdowns, making them attractive options for investors with a low tolerance for risk. RS-RP especially distinguishes itself by efficiently navigating volatility, yielding a Sharpe ratio higher than that of the traditional RP strategy even amidst higher volatility. It effectively alleviates the trade-off between risk management and profit generation. These outcomes showcase

the RS-RP's strategic value for risk-averse investors who still want to achieve a more significant return than that of the traditional RP portfolio.

Meanwhile, we observe that the GSCI index generally contributes to higher volatility for the RP and RS-RP portfolios but also yields higher annualized returns than the DBC index. The portfolios incorporating GSCI tend to have higher Sharpe ratios, indicating the index's capability of yielding higher risk-adjusted returns. However, the DBC index appears to benefit more from both the RP and RS-RP strategies compared to the GSCI index, resulting in higher returns, Sharpe, Calmar, and Sortino values, and in some cases, lower maximum drawdowns for portfolios. Thus, we decide that when an investor aims to exploit the diversification benefits of commodity indices and mitigate risk using the risk parity strategy, either with or without a regime-switching approach, the DBC index is more suitable for such objective. Such strengths are even accentuated when we increase the number and types of assets included in the portfolio, which illuminates the importance of choosing the right set of underlying assets and strategy based on market characteristics and personal investment preferences such as risk-tolerance, investment goals, etc.

Chapter 6

Conclusion

6.1 Summary of Results

In this paper, we aim to explore the benefits of integrating a regime-switching framework into a portfolio with commodity futures. We implement the ℓ_1 trend-filtering algorithm and a continuous statistical jump model for regime identification, and conduct hyperparameter tuning to set the model’s sensitivity to regime changes. Using the regime information obtained from the identification task, we use five models – logistic regression, feedforward neural networks, recurrent neural networks, gradient boosting and random forests – for regime prediction. Throughout this process, we observe the robust performance of logistic regression for regime prediction, especially when paired with the ℓ_1 trend-filtering algorithm, across both GSCI and DBC datasets.

We use this aforementioned model to obtain regime predictions and probabilities, ultimately to backtest three different strategies, mean reversion, trend-following and risk parity, under the regime-switching framework and compare them against traditional portfolios of 60/40 and Equal Weight as well as their counterparts without regime-switching approach. Analysis reveals that the Regime-Switching Mean Rever-

sion (RS-MR) strategy consistently outperforms its Mean Reversion (MR) counterpart and benchmark portfolios, with higher returns as well as superior risk management, both drawdowns and downside risk. The RS-MR offers a compelling option for investors seeking to exploit short-term market inefficiencies.

Similarly, the Regime-Switching Trend-Following (RS-TF) strategy demonstrates its strength with its Sharpe, Calmar, and Sortino ratios exceeding those of traditional trend-following strategies and benchmarks, which can be attributed to the regime-switching component enabling the strategy to dynamically take advantage of observed trends in the market. These findings advocate for the adoption of the regime-switching approach to adapt to prevailing market regimes along with these strategies for portfolio optimization.

The analysis of Regime-Switching Risk Parity (RS-RP) and Risk Parity (RP) requires a more nuanced understanding in an investor's perspective. While these strategies offer lower annualized returns compared to traditional portfolios, they demonstrate their adeptness in better risk-adjustment, suggesting that they are suitable for volatile market environments or investment decisions based on high risk-aversion. RS-RP doesn't consistently outperform the RP across all portfolios, but shows strength in portfolios with higher diversification, which may indicate the necessity to employ with discretion based on specific market conditions.

We also notice that the choice of commodity index and degree of diversification play a critical role in strategy performance due to varying levels of compatibility with the regime-switching approach and market conditions. GSCI and DBC indices show similar levels of compatibility with RS-MR and RS-TF, with the DBC index showing higher compatibility with risk parity strategies in particular. Increased diversification in general leads to reduced downside risk and diluted profits, which particularly comes into play when adopting risk parity strategies.

In summary, these findings highlight the importance of selecting the right strategy,

underlying assets and level of diversification based on market conditions, investment goals, and risk tolerance. The integration of a regime-switching framework into investment strategies presents a promising path for portfolio optimization, opening doors to new opportunities for investors to navigate the complexities of commodity markets, capitalizing on the diversification benefits that accompany the incorporation of such assets into the portfolios.

6.2 Limitations and Future Research

Although this research shows promising outcomes and novel insights, it also has some limitations. One limitation of this research would be that this research does not account for transaction costs, interest-rate environments or liquidity constraints, resulting in a simplified representation of the market dynamics. The actualized efficacy of the strategies that this research focuses on could be affected by the interplay of different factors such as rebalancing frequencies, fiscal policies, risk-free rate, market sentiment, etc. Future research could address these limitations by conducting backtests under diverse scenarios with varying levels of liquidity, rebalancing frequencies, and transaction costs.

Another limitation would be the reliance on specific indices (GSCI and DBC) for the analysis, which could impact the generalizability of the findings across different commodities or asset classes. Commodities can range from precious metals such as gold to energy sources like oil. Each commodity is influenced by a unique combination of factors such as geopolitical events, market dynamics, and technological developments. Such inherent diversity implies that the applicability and effectiveness of our regime-switching framework could vary based on the composition and weights of commodity indices. This raises further question on whether it is better to apply regime-switching at the index versus individual commodity level, given that indices

may dilute individual commodity trends. Future studies could explore the efficacy of these strategies across a broader array of commodities and portfolios, as well as compare the applicability of regime-switching to individual commodity futures.

Lastly, this research allows room for more rigorous testing and implementation of the regime identification and prediction models. This research relies on a two-regime approach, but discretizing the observed data into two regimes may oversimplify subtler shifts that could regardless affect the portfolio performance. Future studies may be able to address this issue through a more continuous and probabilistic approach. Moreover, although the logistic regression model shows solid test accuracy and we aim to mitigate the risk of overfitting through time series cross-validation, there is still room for improvement in terms of prediction model's generalizability to unseen data through further validation, hyperparameter tuning and incorporation of different types of predictive models.

Bibliography

- Alam, M. K., Tabash, M. I., Billah, M., Kumar, S., & Anagreh, S. (2022). The impacts of the russia–ukraine invasion on global markets and commodities: A dynamic connectedness among g7 and bric markets. *Journal of Risk and Financial Management*, 15(8), 352.
- Andersson, H. (2007). Are commodity prices mean reverting? *Applied Financial Economics*, 17(10), 769–783.
- Assoe, K. (1998). Regime-switching in emerging stock market returns. *Multinational Finance Journal*, 2(2), 101–132.
- Aydinhan, A. O., Kolm, P. N., Mulvey, J. M., & Shu, Y. (2023). Continuous statistical jump models for identifying financial regimes. *Available at SSRN 4556048*.
- Bae, G. I., Kim, W. C., & Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, 234(2), 450–458.
- Beckmann, J., Belke, A., & Czudaj, R. (2014). Regime-dependent adjustment in energy spot and futures markets. *Economic Modelling*, 40, 400–409.
- Bemporad, A., Breschi, V., Piga, D., & Boyd, S. P. (2018). Fitting jump models. *Automatica*, 96, 11–21.
- Borgards, O., Czudaj, R. L., & Van Hoang, T. H. (2021). Price overreactions in the commodity futures market: An intraday analysis of the covid-19 pandemic impact. *Resources Policy*, 71, 101966.

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *Biometrics*, 40, 874. <https://api.semanticscholar.org/CorpusID:29458883>
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.
- Cheung, C. S., & Miu, P. (2010). Diversification benefits of commodity futures. *Journal of International Financial Markets, Institutions and Money*, 20(5), 451–474.
- Choi, K., & Hammoudeh, S. (2010). Volatility behavior of oil, industrial commodity and stock markets in a regime-switching environment. *Energy policy*, 38(8), 4388–4399.
- Costa, G., & Kwon, R. H. (2019). Risk parity portfolio optimization under a markov regime-switching framework. *Quantitative Finance*, 19(3), 453–471.
- Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, 62(2), 69–97.
- Fama, E. F., & French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of political Economy*, 96(2), 246–273.
- Fong, S., & Tai, J. (2009). The application of trend following strategies in stock market trading. *2009 Fifth International Joint Conference on INC, IMS and IDC*, 1971–1976.
- Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62(2), 47–68.
- Hamilton, J. D. (2010). Regime switching models. *Macroeconometrics and time series analysis*, 202–209.
- Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A century of evidence on trend-following investing. Available at SSRN 2993026.
- Irwin, S. H., Zulauf, C. R., & Jackson, T. E. (1996). Monte carlo analysis of mean reversion in commodity futures prices. *American Journal of Agricultural Economics*, 78(2), 387–399.

- Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2002). Tactical asset allocation and commodity futures. *Journal of Portfolio Management*, 28(4), 100.
- Kim, M.-J., & Yoo, J.-S. (1995). New index of coincident indicators: A multivariate markov switching factor model approach. *Journal of Monetary Economics*, 36(3), 607–630.
- Kim, S.-J., Koh, K., Boyd, S., & Gorinevsky, D. (2009). \Ell_1 trend filtering. *SIAM review*, 51(2), 339–360.
- Liew, C. C., & Siu, T. K. (2010). A hidden markov regime-switching model for option valuation. *Insurance: Mathematics and Economics*, 47(3), 374–384.
- Maillard, S., Roncalli, T., & Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4), 60–70.
- McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4), 574–589.
- Miffre, J., & Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking & Finance*, 31(6), 1863–1886.
- Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, 104(2), 228–250.
- Mulvey, J. M., Kaul, S. S. N., & Simsek, K. D. (2004). Evaluating a trend-following commodity index for multi-period asset allocation. *The Journal of Alternative Investments*, 7(1), 54–69.
- Mulvey, J. M., & Liu, H. (2016). Identifying economic regimes: Reducing downside risks for university endowments and foundations. *Journal of Portfolio Management*, 43(1), 100.
- Nielsen, S., & Olesen, J. O. (2001). *Regime-switching stock returns and mean reversion* (tech. rep.). Working paper.

- Nystrup, P., Lindström, E., & Madsen, H. (2020). Learning hidden markov models with persistent states by penalizing jumps. *Expert Systems with Applications*, 150, 113307.
- Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of financial economics*, 22(1), 27–59.
- Qian, E. (2011). Risk parity and diversification. *The Journal of Investing*, 20(1), 119–127.
- Szakmary, A. C., Shen, Q., & Sharma, S. C. (2010). Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking & Finance*, 34(2), 409–426.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Uysal, A. S., & Mulvey, J. M. (2021). A machine learning approach in regime-switching risk parity portfolios. *The Journal of Financial Data Science*.
- Wang, Y., Bouri, E., Fareed, Z., & Dai, Y. (2022). Geopolitical risk and the systemic risk in the commodity markets under the war in ukraine. *Finance Research Letters*, 49, 103066.
- Yamada, H. (2018). A new method for specifying the tuning parameter of 1 trend filtering. *Studies in Nonlinear Dynamics & Econometrics*, 22(4), 20160073.
- Zhang, Y., & Wang, R. (2022). Covid-19 impact on commodity futures volatilities. *Finance Research Letters*, 47, 102624.