GRADUATE SCHOOL OF NATURAL SCIENCES

COMPUTING SCIENCE

Top-Down and Bottom-Up replication of Commodity Trading Advisors

Author: Kan Li

Student Number: 2216612

Supervisor: Dr. Matthijs Vákár

Second Reader: Dr. Thomas Walther

Date: July 15, 2024

Abstract

This research aims to replicate CTAs using both top-down and bottom-up methods. The top-down method involves estimating CTA positions based on their daily returns over some lookback period, by using linear regression. The bottom-up method combines trend and carry strategies, followed by an optimization process to determine the best weights for strategy variants. The robustness of these methods is evaluated through in-sample and out-sample tests. Performance metrics such as R^2 and Sharpe ratio are used to assess the methods. The results indicate that the bottom-up method performs well in replication, with a Sharpe ratio of 0.8 and an R^2 of 0.6, also demonstrating good robustness. While the top-down method achieves a Sharpe ratio of 0.47 and an R^2 of 0.42, also demonstrating poor robustness. In terms of accuracy, profitability, and robustness, the bottom-up method outperforms the top-down method.

Contents

1 Introduction **4**

2 Research background **4**
2.1 CTA replication . . . 4
2.1.1 Why is CTA replication hard? . . . 4
2.1.2 What does it mean to be successful? . . . 5
2.1.3 What has been done . . . 5
2.1.4 ReSolve Asset Management . . . 6
2.1.5 DBMF . . . 9
2.2 Research questions . . . 9
2.3 Methods and data . . . 10
2.3.1 Carry . . . 10
2.3.2 Trend-following . . . 11
2.3.3 Data . . . 12
2.4 Evaluation metrics . . . 13
2.4.1 Sharpe ratio . . . 13
2.4.2 R^2 . . . 14
2.5 Risk management . . . 14

3 The top-down method **15**
3.1 Procedural framework . . . 15
3.2 Heatmap and correlation analysis . . . 15
3.3 Selection of instruments . . . 16
3.4 Rolling regression . . . 16
3.4.1 Regularization methods . . . 19
3.5 Flow of the top-down method . . . 20

4 The bottom-up method **22**
4.1 Risk management . . . 22
4.1.1 Measuring risk . . . 22
4.1.2 Position scaling . . . 23
4.2 Trend . . . 24
4.2.1 Trend forecast . . . 24
4.2.2 Flow of Trend . . . 25
4.3 Carry . . . 27
4.3.1 Measuring Carry . . . 27
4.3.2 Carry forecast . . . 27
4.3.3 Flow of Carry . . . 28
4.4 Optimization . . . 30
4.4.1 Objective functions and constraints . . . 30
4.4.2 Flow of optimization . . . 31
4.4.3 Overfitting . . . 31
4.5 Conclusions . . . 31

5 Results and analysis **33**
5.1 Results of top-down methods . . . 33
5.2 Robustness of the top-down method . . . 35
5.3 Results of bottom-up methods . . . 39
5.4 Robustness of the bottom-up method . . . 41
5.5 Comparison and analysis . . . 47

6 Limitations and Future work **48**
6.1 Limitations . . . 48
6.1.1 Trading costs and buffer . . . 48
6.1.2 Scope of the bottom-up method . . . 48
6.1.3 Quantification of Uncertainty in Strategy Reconstructions . . . 48
6.2 Future work . . . 49
6.2.1 Extreme values in the bottom-up method . . . 49
6.2.2 Differences between the results of the top-down method and DBMF . . . 49
6.2.3 Combination of the bottom-up and top-down methods . . . 49
6.2.4 Investigation of Long-Only Biases . . . 49

1 Introduction

Investment funds known as Commodity Trading Advisors (CTAs) present intriguing investment prospects due to their diversification attributes compared to the predominant long-only buy and hold equity portfolios commonly held by individuals and institutions. However, CTAs are encumbered by their high fees and the inherent opaqueness associated with being classified as black box funds, thereby impeding potential investors. Another motivation is that there exists a considerable amount of misconception surrounding the role and actions of CTAs. An example showing this misconception is an article "Hedge funds profit from Ukraine war food price surge" published by The Guardian [Guardian, 2023]. The article suggests that it is unethical for CTAs to profit from the Ukraine war. However, it fails to acknowledge that CTAs had already established substantial positions in wheat prior to the onset of the conflict. To address these issues, this project aims to employ statistical modeling techniques to unveil the inner workings of CTAs, thereby effectively unlocking the black box and reconstructing their internal positions exclusively utilizing publicly accessible performance data.

This research aims to shed light on the strategies and composition of CTAs, offering valuable insights for investors to make informed decisions. Beyond benefiting individual investors, this study provides governmental bodies and regulatory agencies with critical insights to better supervise and manage CTA operations, thanks to increased transparency in their strategies and compositions. Understanding CTAs helps evaluate the potential risks and impacts associated with these funds, serving as a crucial reference for policy and regulation formulation. These measures ensure the healthy development of the market and protect investor interests.

2 Research background

2.1 CTA replication

2.1.1 Why is CTA replication hard?

The replication of CTAs presents considerable challenges due to various factors such as the scarcity of data, the vagueness of what strategies that CTAs are running, and the ambiguity of evaluation metrics. This section delves into the reasons behind the difficulty in accurately replicating CTAs and highlights key issues associated with data availability, success criteria, the elusive nature of CTA strategies, and the limited academic research in this domain.

Data perspective CTAs, like most other hedge funds, tend not to disclose their positions and publish, at most, daily, weekly, or monthly return data. This limited transparency presents significant challenges for replication. Without detailed position data, we cannot see the specific assets and strategies employed, making it difficult to understand the performance drivers accurately. The sparse return data available lacks the granularity needed for thorough analysis, hindering the capture of dynamic strategy adjustments.

Vaguesness of what strategies CTAs are running Furthermore, the elusive nature of CTA strategies poses a significant hurdle in the replication process. The majority of CTAs do not disclose their strategies, and even if they do, there is no guarantee that the publicly disclosed strategies reflect their actual practices. This opacity inhibits replication efforts, as it becomes arduous to accurately discern and replicate the intricate decision-making processes and trading strategies employed by CTAs.

2.1.2 What does it mean to be successful?

The challenge of replicating CTAs is made harder by the difficulty in defining success criteria. To properly assess replication models, we need metrics that match the goals and performance measures of CTAs. However, setting clear benchmarks is tough due to the wide range of strategies and investment goals used by CTAs. Therefore, agreeing on the right metrics to evaluate replication success is still a challenge in both academia and industry.

Additionally, it is important to focus on the main goal of CTA replication. While exactly matching post-fee returns is very difficult, especially considering the fees charged by these funds, the main aim might be to create a profitable investment strategy. So, using risk-adjusted return measures, like the Sharpe ratio, Sortino ratio, and Calmar ratio, is a wise approach to assess the profitability of the replicated strategy. In summary, success in CTA replication should be judged by both how accurately returns are replicated and how profitable the strategy is, in risk-adjusted terms, compared to the target CTA.

2.1.3 What has been done

According to Takahashi et al. [2008], the biggest banks such as Goldman Sachs, Merrill Lynch, and JP Morgan, and some large investment companies such as, Partners Group, have launched hedge fund replication products. There are 3 hedge fund replication methods: Rule-based, Factor-based, and Distribution replicating approaches. First, the Rule-Based Replication method involves replicating hedge fund strategies by following a set of predetermined rules or algorithms that mimic the decision-making processes of fund managers. This approach aims for transparency and simplicity but may struggle to capture the nuanced strategies of active fund managers. Second, the Factor-Based Replication method uses regression analysis to extract a set of factors from public market data, establishing a linear model to replicate hedge fund returns. This method is transparent, easy to understand and implement, and cost-effective, but it is sensitive to factor selection and may not fully capture the unique skills of fund managers. Finally, the Distribution Replicating Approach analyzes historical data to understand the distribution of hedge fund returns, using statistical methods to simulate and replicate this distribution. This approach better captures the risk and return characteristics of the fund but requires extensive historical data and assumes future performance will mirror the past.

Amenc et al. [2010] examines the out-of-sample performance of diverse non-linear and conditional models used for replicating hedge funds. Their investigation reveals that venturing beyond the linear framework such as linear regression does not necessarily enhance the effectiveness of replication. However, they observe a noteworthy improvement in out-of-sample replication quality when factors are selected based on an economic analysis, regardless of the underlying form of the factor model. In line with the findings of Hasanhodzic and Hasanhodzic and Lo [2006], their overall results confirm that the replicating strategies consistently underperform actual hedge funds.

Fung and Hsieh [1997] focused on analyzing the impact of survivorship bias in private funds managed by CTAs. The study concluded that CTA funds dissolve more frequently than mutual funds and that the dissolution cost is considerably higher. It also observed that CTAs operating multiple funds tend to behave differently in terms of risk management to protect their reputation, particularly when a fund is underperforming. The dominant investment style in both surviving and dissolved CTA funds was identified as a trend-following strategy, which the researchers were able to measure with minimal survivorship bias. This strategy had a low correlation to standard asset classes, suggesting its unique risk-return profile within the spectrum of investment strategies

Gregoriou and Chen [2006] examined the performance of CTAs using fixed and variable benchmarking models. To circumvent the challenges posed by passive and active commodity and managed futures benchmarks (indices) in evaluating CTA performance, the authors innovated by utilizing data envelopment analysis (DEA). Given the non-linear returns characteristic of this alternative asset class, attributed to long/short positions and derivatives, DEA was employed to mitigate the issues commonly associated with these indices. The efficacy of benchmarking models within a DEA framework provided investors with an alternative method for assessing performance and identifying efficient CTAs.

Bhardwaj et al. [2014] analyzed the performance of CTAs from 1994 to 2012. They evaluated the performance of CTAs while accounting for various biases like strategic reporting and database construction issues. They found that after adjusting for biases, the average CTA's returns, net of fees, were essentially the same as U.S. Treasury bills, implying that CTAs did not generate significant value for investors. This included a newly identified "graveyard bias" where entire fund records were removed from databases, affecting performance analysis.

2.1.4 ReSolve Asset Management

In recent marketing materials conducted by ReSolve Asset Management, the authors investigate two distinct methods for the replication of CTA strategies: top-down replication and bottom-up replication. The top-down approach involves an analysis of the returns exhibited by the target index during a defined period, enabling the determination of the average holdings maintained by the managers included in the index at each specific point in time. Conversely, the bottom-up approach seeks to uncover the fundamental strategies employed by the funds encompassed within the index, thereby facilitating the construction of their respective portfolios. By gaining insight into the weighting and composition of representative strategies, the replication strategy can emulate the holdings of the index over a given time frame. Consequently, the bottom-up approach relies on an informed estimation of the weights assigned to different instruments and trading signals based on the perceived underlying strategies employed by the managers.

The authors highlight the challenges in replicating trend-following strategies due to the dispersion of returns among individual funds. While the Barclays BTOP Index and similar indices provide a broad representation of trend-following funds, allocating to a single fund or a small sample of funds may not capture the category-like returns consistently. This is because funds can differ in their choice of markets, trend lengths, weighting schemes, risk management techniques, and exposure to other systematic factors. As a result, returns among trend-following managed futures funds can vary significantly from year to year.

To mitigate dispersion and attain exposure to the trend-following theme while minimizing tracking error, the utilization of replication strategies becomes imperative. Such strategies aim to closely replicate the return trajectory of a designated benchmark. In the context of trend-following managed futures, the authors of the study delve into the challenges associated with selecting an appropriate benchmark for replication. Although the widely adopted Barclay BTOP 50 Index is commonly employed, its drawback lies in the availability of daily returns data only from 2010 onwards, which proves suboptimal for replication endeavors. As an alternative, the authors propose the SG CTA index, which tracks the ten largest trend-following managed futures funds and provides daily returns data dating back to January 2000. The utilization of this index offers a more comprehensive and extensive dataset, facilitating a robust assessment of replication effectiveness. Moreover, it would be intriguing to explore the divergence between the replication of the BTOP 50 Index and the SG CTA Index, providing insights into the dissimilarities in replication outcomes between the two benchmarks.

Due to the lack of access to the actual holdings and trading data of the funds in the benchmark, replication strategies must infer the weighted average holdings based on an understanding of the markets typically traded by managed futures funds, the strategies they employ, and reported index returns. This publication uses two methods to replicate CTAs: top-down and bottom-up. The authors use data from Société Générale Prime Services as the Daily returns of the Trend Index and CSI Data and Refinitiv as the futures market data.

Top-down method The top-down method involves replicating a portfolio by regressing the track record of an individual or index on a universe of stocks or assets to identify the optimal combination that closely matches the returns.

In the top-down approach, the authors utilized Elastic Net regression for linear regression analysis. This method aimed to identify the best replication portfolio that mimics the returns of CTAs by examining the relationship between the historical performance of CTAs and the returns of various markets.

On a daily basis, the model was fit based on the past few days or weeks of index and explanatory market returns, yielding weights for representative markets. These weights minimized the squared differences between the tracking portfolio's returns and the benchmark returns during the regression period. Elastic Net regression was chosen because it could identify the most influential markets in each period while penalizing models that were overly concentrated in any single market.

Rolling returns were used to fit the Elastic Net model, incorporating past 20, 25, 30, 35, and 40 days of returns for each fit. The tradeoff in replicating index returns lies between the breadth of markets used and the model fit. A larger market universe explains a greater proportion of returns but compromises reliability when the number of explanatory variables outweighs the available observations. A smaller universe may not capture the full opportunity set but yields more stable results with limited observations. Ultimately, the linear model was fit to the model outputs of each backtest period to determine the weights for the final model.

Therefore, the authors chose two representative market combinations, one smaller portfolio consisting of 9 markets representing broad exposures to stocks, fixed income, commodities, and currencies, and another larger portfolio consisting of 27 markets with more refined exposure within the same industries. When fitting the weights using Elastic Net regression, both market combinations produced similar-quality fits and performance over time in each period. However, by combining the models, a better tracking portfolio could be generated than with individual models, particularly in periods of divergent performance.

Bottom-up method The bottom-up method focuses on identifying and replicating the underlying strategies or characteristics used by funds or managers to select individual stocks or assets. In the bottom-up approach, the authors employed a range of trend-following trading strategies to identify market trends. Thirteen representative trend-following strategies were selected, and their calculations were performed using different lookback periods. The trend for each strategy was measured by the z-score derived from the ratio of total returns to the historical standard deviation of the index-weighted past 40 days.

The position of each market in each trend strategy was determined by multiplying the z-score of that market by the expected volatility based on the index-weighted historical standard deviation. This resulted in a combination of 13 trend strategies multiplied by 27 markets, yielding a total of 351 market/strategy pairs.

Subsequently, the authors employed Ridge regression to fit the returns of these 351 strategies, aiming to minimize the differences between the replicated portfolio's returns and the index returns when trading according to the model-specified weights. The model weights were constrained to be positive values only. Ridge regression was chosen over LASSO or

Elastic Net because it is well-suited for identifying all relevant predictive factors, even if some may be falsely identified.

As the strategies employed in trend-following funds have a strong record in the literature, the focus was on ensuring the inclusion of all relevant models, irrespective of their potential underutilization by trend-following managers. K-fold cross-validation was used to assess if strategy exposures changed over time. The model was fitted on 90% of the data, and the selected strategies were run on the remaining 10% data. The out-of-sample strategy returns were then combined, simulating the performance of the complete history of the SG Trend Index. Additionally, the model was fitted on the complete data sample, and the selected strategies were run on the full dataset to generate a full simulation based on in-sample fitting.

To compare the performance of the two simulations, the daily returns were scaled to the same target volatility using a rolling 40-day index-weighted standard deviation. The out-of-sample simulation demonstrated comparable performance to the simulation conducted on the entire dataset, exhibiting a high daily Pearson correlation coefficient of 0.997. This finding suggests that the model fitted using the complete sample data is capable of accurately tracking the index in future periods. Furthermore, the selected strategies remained consistent across various sub-samples, indicating that trend-following futures managers maintained stability in their strategies and core market selections over time.

Discussion of the bottom-up methods by ReSolve Asset Management The bottom-up replication method discussed above focuses on the identification and replication of underlying strategies or characteristics utilized by funds or managers in the selection of individual stocks or assets. However, it is important to acknowledge potential areas for improvement in the method, as suggested in the provided comment.

One pertinent critique is that the current approach neglects to account for the tendency of CTAs to refrain from altering their positions overnight. To address this concern, the integration of a modeling technique such as the Kalman filter or a similar model, capable of explicitly capturing the interdependence between regression results across consecutive days, could prove advantageous. While the existing method benefits from the utilization of overlapping data, explicitly modeling this relationship would offer additional insights.

Moreover, there is a substantial issue of trading costs associated with rapid position changes. To mitigate this concern, the incorporation of a penalty term for trading costs within the optimization process presents a potential solution. By incorporating the impact of trading costs into the optimization framework, the replication strategy could more accurately reflect the real-world constraints faced by CTAs. However, it is crucial to implement this adjustment unless it can be substantiated that the trading costs already remain at a sufficiently low level.

Final strategies The top-down approach in replication has the advantage of identifying the most important markets during each period, allowing for a focused allocation of resources. However, it may exhibit a slightly inferior fit compared to the bottom-up method. On the other hand, the bottom-up approach excels in achieving a superior fit and overall performance, but raises concerns about the potential for new and unrepresented strategies implemented by future managers. Therefore, while the top-down method displays a lower tracking error during crucial times in historical data, the bottom-up method provides a better fit. Combining both approaches harnesses their respective advantages, yielding higher performance than either approach individually and improving the long-term correlation between replication and benchmark returns.

The final replication strategy consists of a combination of top-down and bottom-up approaches, with a 50% weight given to each method. This strategy is chosen because it effec-

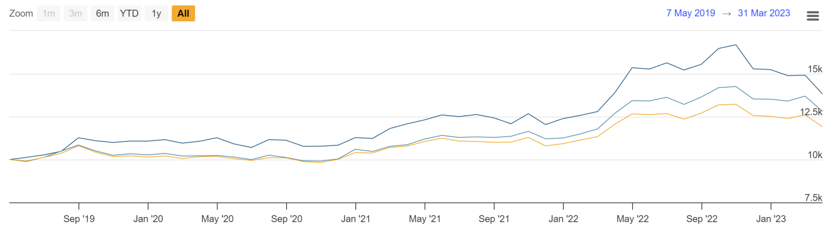

Figure 1: Performance of DBMF [2013].

tively balances the strengths and weaknesses of both approaches, creating a tracking portfolio that is accurate and stable. By incorporating both top-down and bottom-up methods, the strategy captures the broad exposures to equities, fixed income, commodities, and currencies present in the benchmark index, while also identifying the underlying strategies used by trend-following managed futures funds.

2.1.5 DBMF

The iMGP DBi Managed Futures Strategy ETF (DBMF) utilizes a top-down approach to replicate CTAs. DBMF [2013] shows they employ an index-based investment strategy that aims to generate superior risk-adjusted returns through transparent quantitative techniques and rules-based criteria based on specific factors or attributes that drive investment returns. Their portfolio holdings include various assets such as Treasury bills, euro foreign exchange, gold, Japanese Yen, and US long bonds, among others.

The strategy employed by DBMF has shown strong performance. According to Figure 1, the fund's returns have consistently outperformed both the SG CTA index and the Morningstar US Systematic Trend Category since its inception. This demonstrates the viability of DBMF's CTA replication approach and provides empirical support for this research.

However, the specific statistical models or quantitative techniques used by DBMF are not disclosed in the information provided. This lack of transparency regarding their strategies further motivates this research to delve into uncovering the industry secrets of CTAs. By exploring and understanding the undisclosed method and techniques utilized by successful CTAs like DBMF, this research aims to shed light on the strategies employed within the CTA industry.

2.2 Research questions

This study will use top-down and bottom-up approaches to replicate CTAs. Top-down approach involves using regression models to calculate positions of managed futures. The bottom-up approach involves using trend-following and carry strategies to identify the trading strategies or characteristics of fund managers. Based on the related background and previous work of CTA, this section poses the following research questions:

- Top-down approaches: Can the top-down approach effectively replicate the positions of CTAs? If so, how much can the top-down approach explain? How does the effectiveness vary with the number of selected instruments? Are there other parameters that influence the results, such as the size of the rolling window? Would regularization be beneficial for replication? If so, which regularization method is more suitable? Is the top-down approach profitable? What metrics can be used to assess success? How is the robustness of the top-down method? How can we best evaluate the robustness of the top-down method?

- Bottom-up approaches: Can the bottom-up approach provide a detailed analysis of the trading strategies used by CTAs, especially for carry and trend-following strategies? Is it possible to replicate CTAs using the bottom-up method? Can optimization methods be employed to determine the optimal allocation of carry and trend-following strategy signals? If so, what proportion of signals should be dedicated to carry and trend-following strategies? What is the distribution of signals across various asset classes? Can optimization methods be employed to determine the optimal allocation of carry and trend-following strategy signals? If so, what proportion of signals should be dedicated to each asset class? Is the bottom-up method robust? How can we evaluate its robustness? In comparison to the top-down approach, does the bottom-up method offer higher accuracy and greater profitability?

2.3 Methods and data

2.3.1 Carry

Ralph S.J. Koijen et al. explain carry in detail and examine carry in many different asset classes in Koijen et al. [2018]. In their study, the concept of "carry" is introduced as the futures return of an asset under the assumption of constant prices. The excess return can be expressed as the sum of the carry and spot return components:

$$\text{Excess return} = \text{Carry} + \text{Spot return}$$

Carry, which represents a model-free characteristic of an asset and can be directly observed from futures prices, is distinct from expected price appreciation, which requires estimation using asset pricing models. The authors empirically examine carry across various asset classes and demonstrate its relationship with expected returns and price appreciation.

The article explores the predictive power of carry in different asset classes and its ability to test asset pricing theories. The findings reveal that carry is a strong positive predictor of returns in each asset class studied, both across the cross-section and over time. In their analyses, a carry trading strategy involves long positions in high-carry assets and short positions in low-carry assets, generated significant returns with a Sharpe ratio of 0.8 on average, on the time period examined (1983-2012). Furthermore, a diversified portfolio of carry strategies across all asset classes exhibits a Sharpe ratio of 1.2. The returns from carry cannot be explained by other known return predictors, indicating its unique contribution to return predictability.

The study highlights the unifying nature of carry as a predictor of returns across different asset classes, suggesting that it could serve as a unifying concept that connects various return predictors scattered across the literature. By examining carry in conjunction with other known factors, such as value and momentum, the authors demonstrate its distinct predictive power. They also investigate potential explanations for the carry premium, including crash risk and downside risk exposure, but find that these factors alone cannot fully account for the observed returns.

Carry is more closely associated with specific asset classes. Various financial instruments within different asset classes derive their carry from diverse sources. For instance, in the case of equities, carry is influenced by factors such as dividend yields and interest rates. This relationship can be mathematically expressed as follows, according to Carver [2023b]:

$$\text{Excess return(equities)} = \text{Spot return} + \text{Dividend} - \text{Interest}$$

$$\text{Carry(equities)} = \text{Dividend} - \text{Interest}$$

In this context, the interest rate is specified as the risk-free rate, and the excess return is the return over the risk-free rate.

2.3.2 Trend-following

A trend-following strategy entails the adoption of long or short positions based on the prevailing market trend. Its profitability is contingent upon the presence of sustainable price trends in future contracts. If the price of a particular market has exhibited an upward trajectory over the past year, a trend strategy anticipates a continuation of this upward trend, while in the case of falling prices, a further decline is expected. The primary objective of Trend Following is to capture momentum by utilizing the historical returns of macro asset classes.

Since the last century, it has been discovered that it is possible to profit through momentum. In 1999, Jegadeesh and Titman [2001] evaluated the profitability of momentum strategies based on Jegadeesh and Jegadeesh and Titman [1993] findings. Using an extended dataset from 1990 to 1998, the authors confirmed that momentum strategies remained profitable, supporting the original results. They found evidence in favor of behavioral models, suggesting that momentum profits may result from investor biases and delayed reactions to information. However, they cautioned against overinterpretation and highlighted the ongoing debate about the interpretation of momentum profits. The study emphasized the need for further research to understand the underlying causes of momentum profits and their implications for investment strategies and market efficiency.

Cross-sectional momentum and time-series momentum are commonly used momentums by traders and investors. Cross-sectional momontum was first proposed by Jegadeesh and Titman [2002] in 2002. Cross-sectional momentum focuses on the relative performance of different assets within a specific period. It compares the performance of various assets, such as stocks or sectors, at a given point in time. The idea is to identify the assets that have exhibited the strongest relative performance compared to their peers. Cross-sectional momentum typically involves ranking assets based on their recent performance and constructing a portfolio that includes the top-performing assets while excluding the underperforming ones.

Time-series momentum was prosed by Moskowitz et al. [2012] from AQR in a paper titled "Time Series Momentum" in the Journal of Financial Economics. The study investigated the momentum effect of various asset classes across the globe over a span of 25 years. The findings revealed a significant presence of "time series momentum effects" in the prices of 58 financial assets, including stock index futures, foreign exchange forwards, commodity futures, and bond futures. This momentum effect is not an average outcome across the 58 assets but rather manifests as positive returns on each asset. These returns exhibit persistence in the short term (1 to 12 months) and reversal in the long term (beyond 12 months).

Compared to cross-sectional momentum, time-series momentum or trend following focuses on the historical performance of a single asset over time. It aims to capture the persistent price trends exhibited by individual assets. Time-series momentum traders typically look for assets that have shown consistent price increases or decreases over a specified time horizon. They enter long positions when the asset price has been trending upward and short positions when the price has been trending downward.

In 2013, Asness et al. [2013] provided evidence that value and momentum are widespread across asset classes. This academic study explores the value and momentum return premia across eight diverse markets and asset classes. The researchers make significant contributions to the existing literature by discovering consistent value and momentum returns in all the examined markets, including government bonds, currencies, and commodities. A notable finding is the presence of a robust common factor structure among the returns of value and momentum strategies across various asset classes. While positive correlations are observed within each strategy category, negative correlations are identified between value and momentum both within and across asset classes. To explain these patterns, the researchers propose a three-factor model that effectively captures global returns across asset classes, exhibiting explanatory power for US stock portfolios and hedge fund indices. The study further reveals global funding liquidity risk as a partial underlying factor for the observed phenomena. This empirical evidence challenges prevailing behavioral, institutional, and rational asset pricing theories. Importantly, the study underscores the necessity of jointly examining value and momentum, emphasizing their interaction and common factor structure as crucial elements in comprehending market dynamics.

The Exponentially Weighted Moving Average (EWMA) is a tool commonly used in technical analysis to smooth time series data. We use EWMA in the bottom-up method to catch the trends of futures prices. Using the span parameter, the calculation process can be more intuitively understood. The relationship between the span and the smoothing factor α is as follows:

$$\alpha = \frac{2}{\text{span} + 1} \tag{1}$$

Thus, the EWMA can be calculated using the formula:

$$\text{EWMA}_t = \alpha \cdot x_t + (1 - \alpha) \cdot \text{EWMA}_{t-1} \tag{2}$$

where:

- EWMA_t is the Exponentially Weighted Moving Average at time t.
- x_t is the observed value (e.g., price) at time t.
- α is the smoothing factor, calculated as $\frac{2}{\text{span}+1}$.
- EWMA_{t-1} is the Exponentially Weighted Moving Average at the previous time period $t - 1$.

The initial value EWMA_0 is typically set to the first observed value x_0:

$$\text{EWMA}_0 = x_0 \tag{3}$$

Combining the above, the EWMA calculation can be summarized as follows:

$$\text{EWMA}_t = \frac{2}{\text{span} + 1} \cdot x_t + \left(1 - \frac{2}{\text{span} + 1}\right) \cdot \text{EWMA}_{t-1} \tag{4}$$

with the initial condition:

$$\text{EWMA}_0 = x_0 \tag{5}$$

2.3.3 Data

The data in this work are divided into two main parts. The first part is about the return data of CTAs, shown by Figure 3. This work uses data from PLC. [2023] as target return to replicate. The second part includes price data for futures. For future price data, this work uses data from Robert Carver's pre-possessed data Carver [2023b]. The futures price data consists of two essential parts: the time when the data was recorded and the corresponding prices. Figure 2 is a preview of the first few rows of aluminum price data. Preparing the futures data involves the following steps:

	A	B
1	DATETIME	price
2	2014-06-02 23:00:0C	2531.25
3	2014-06-03 23:00:0C	2535
4	2014-06-04 23:00:0C	2526.5
5	2014-06-05 23:00:0C	2538.5
6	2014-06-06 23:00:0C	2561

Figure 2: Sample price data.

	A	B	C	D	E	F
1	Dstamp	ROR	VAMI	MTD	QTD	YTD
2	1/3/2000	-0.33%	996.74	-0.33%	-0.33%	-0.33%
3	1/4/2000	-0.98%	986.9919	-1.30%	-1.30%	-1.30%
4	1/5/2000	-0.29%	984.1592	-1.58%	-1.58%	-1.58%

Figure 3: Sample return data.

1. Removal of rows with missing prices, along with standardization of date formats.

2. Removal of time-of-day information, thereby retaining only the date component. In scenarios where multiple price data entries correspond to the same date, the average value of these entries is computed.

3. Given that our regression's independent variable hinges on ratios of return, it becomes essential to transform the daily prices into relative increase ratios concerning the previous day. This transformation is explained by the following formula:

$$\text{Relative Increase Ratio} = \frac{\text{Backadjusted Price}(n)}{\text{Raw Futures Price}(n-1)} - 1$$

   Where:

   - Backadjusted Price(n) denotes the backadjusted price of the commodity on the nth day.
   - Raw Futures Price$(n-1)$ signifies the raw price of the commodity on the $(n-1)$th day.

4. Conversion of percentages into decimals.

Raw futures prices are the actual prices at which futures contracts are traded in the market. Backadjusted prices, on the other hand, are adjusted prices that account for contract rolls and other changes over time. These adjustments help to create a continuous price series that is more suitable for analysis. Backadjusted prices remove gaps caused by contract expirations and the introduction of new contracts, providing a smoother and more consistent historical price series. Using backadjusted prices for calculating returns provides a more accurate reflection of the underlying trends and movements in the market by eliminating distortions caused by contract rolls.

2.4 Evaluation metrics

2.4.1 Sharpe ratio

The Sharpe ratio is a measure of risk-adjusted return that evaluates the excess return earned per unit of risk taken. It is defined as the ratio of the portfolio's average excess return to its standard deviation of returns.

Formula:

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p}$$

where R_p is the portfolio's average return, R_f is the risk-free rate, and σ_p is the standard deviation of the portfolio's returns.

2.4.2 R^2

R^2 is the percentage of the response variable variation that is explained by a linear model. It is always between 0 and 100%. We are going to use R^2 a metric because it indicates how well the model accounts for the changes in the dependent variable. A higher R^2 value indicates a better fit, meaning the model accounts for a larger proportion of the variation, which helps in assessing the model's predictive accuracy and overall performance. The formula for R^2 is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where:

$$\begin{aligned} y_i &: \text{Actual values of the dependent variable,} \\ \hat{y}_i &: \text{Predicted values of the dependent variable,} \\ \bar{y} &: \text{Mean of the actual values,} \\ n &: \text{Number of observations.} \end{aligned}$$

2.5 Risk management

In the context of CTAs, allocating position weights based on risk parity is a common practice. This approach focuses on considering the risk associated with different assets or positions, rather than solely relying on prices.

The rationale behind allocating weights based on risk parity is to ensure that the portfolio is balanced and diversified, taking into account the varying levels of risk across different assets. By considering risk, rather than just prices, investors aim to achieve a more efficient allocation of capital and manage risk exposure effectively.

To allocate position weights based on risk parity in a portfolio, the following formula can be utilized:

$$\omega_i = \frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j} \quad (6)$$

In this formula: - ω_i represents the weight allocated to asset or position i. - α_i denotes the risk measure associated with asset or position i. - The denominator, $\sum_{j=1}^{N} \alpha_j$, represents the sum of risk measures for all N assets or positions in the portfolio.

The weight assigned to each asset or position is determined by dividing its individual risk measure by the sum of risk measures across all assets or positions in the portfolio. This ensures that the weights are proportional to the relative risk levels of each asset or position.

By allocating weights based on risk parity using the Greek letters ω and α, the portfolio construction aims to achieve a balanced and diversified allocation, considering the varying levels of risk across different assets or positions.

3 The top-down method

The last section mentioned that DBMF has successfully replicated the performance of leading managed futures hedge funds and, in some cases, even surpassed them by strategically eliminating fees and expenses. However, the detailed method and statistical models used for this replication have not been openly shared. This lack of detail has led to significant interest in understanding and possibly replicating these successes.

This section aims to clarify how one replicate the ratios of return of CTAs using the top-down method. CTAs often trade at regular intervals and tend to distribute their investments evenly across various trading instruments. A rolling regression method, therefore, could be suitable for replicating their return ratios. The method is to use linear regressions within a lookback period, using CTA's daily return as training data to predict the next day's ratio of return.

3.1 Procedural framework

In this section, we will explain the steps involved in top-down method. We start by preparing the raw dataset. This preparation includes fixing missing data, converting data types, removing price data of instruments with too few samples, and selecting the relevant time periods for price and return ratio data. We then convert daily price data into daily increment ratios to show the relative increase from the previous day's values.

Next, we use the increment ratio data to create a heatmap for each instrument. This heatmap shows how different instruments relate to and affect each other. It helps us understand the market dynamics that influence these instruments.

After creating the heatmap, we carefully select which instruments to include in our analysis. We start by randomly picking the first instrument. Then, we add more instruments one by one, choosing each new instrument to lower the average correlation with those already selected. By repeating this process, we create a diverse group of n instruments that represent various aspects of the market.

We then use a rolling regression method to model how these instruments comprise the portfolio. The method involves using linear regressions over a lookback period, with CTA's daily returns as training data to predict the next day's return ratio. Key parameters include the rolling time window size (which determines the amount of training data for each regression), the number of instruments, and the regularization method (ElasticNet, Lasso, Ridge).

From the rolling regression results, we calculate R^2 and the Sharpe ratio as performance metrics. The metrics can comprehensively evaluate the performance of the model and strategy. R^2 provides a measure of the model's goodness of fit. A higher R^2 means a better explanatory power of the model. The Sharpe ratio offers an assessment of risk-adjusted returns.

To explore how parameters influence the method's performance, we conduct a series of tests with different values of parameters, including the window size, the number of instruments, and the regularization methods. To test the robustness of the method, we use other CTAs as target return to replicate their cumulative returns.

3.2 Heatmap and correlation analysis

In our study, we examined the relationships between different instruments. To simplify the understanding of these relationships, we used heatmaps. These heatmaps illustrate the degree of connection or dependence between the instruments. This involved calculating how closely the movements of one instrument are related to another's. The metric we used is the correlation coefficient, which quantifies the degree of relationship between two variables. These coefficients capture the linear relationships between the variables, offering insights into the strength of their associations. The formula used to determine this correlation coefficient is as follows:

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_j (Y_j - \overline{Y})^2}} \tag{7}$$

where:

- X_i and Y_i are individual sample points.
- $\overline{X}$ and $\overline{Y}$ are the means of the X and Y samples, respectively.
- Y_j represents individual sample points in the Y sample.
- The summations in the numerator and denominator are taken over all sample points.

To visually represent these correlations, we have generated a heatmap as shown in Figure 4. The heatmap provides an intuitive depiction of the intricate web of relationships among the instruments. Variables with lower correlation coefficients are preferred choices for inclusion in our regression analysis due to their lower interdependence. If the independent variables have low correlation with each other, it is easier to identify and interpret the individual impact of each variable on the dependent variable. This aids in understanding the relationships between each independent variable and the dependent variable more clearly.

3.3 Selection of instruments

To make it easier to select any number of instruments for future experiments, this study uses an automated method that picks instruments based on their average minimum correlation. The procedure unfolds as follows:

1. Randomly select an instrument.
2. From the remaining instruments, pick another instrument that minimizes the average correlation with the instruments already chosen.
3. Repeat step 2 until the desired count of n instruments has been selected.

This method ensures that the chosen instruments are diversified and possess a lower average correlation, thus enhancing the robustness of subsequent analyses.

Algorithm 1 provides a pseudocode representation of the instrument selection process:

The algorithm results show that it tends to select instruments from diverse asset classes, ensuring a varied set covering different market sectors. Throughout experiments, we observed that though the choice of instruments varied with different parameter settings, the algorithm consistently produced diverse sets of instruments. This indicates that our method adapts to various market conditions and captures correlations and dynamics across different asset categories.

3.4 Rolling regression

CTAs typically engage in daily trading activities, but their positions are often maintained for weeks or months due to the relatively slow nature of their trading signals. This slower pace of trading signal changes is a key characteristic. Given this context, the top-down method

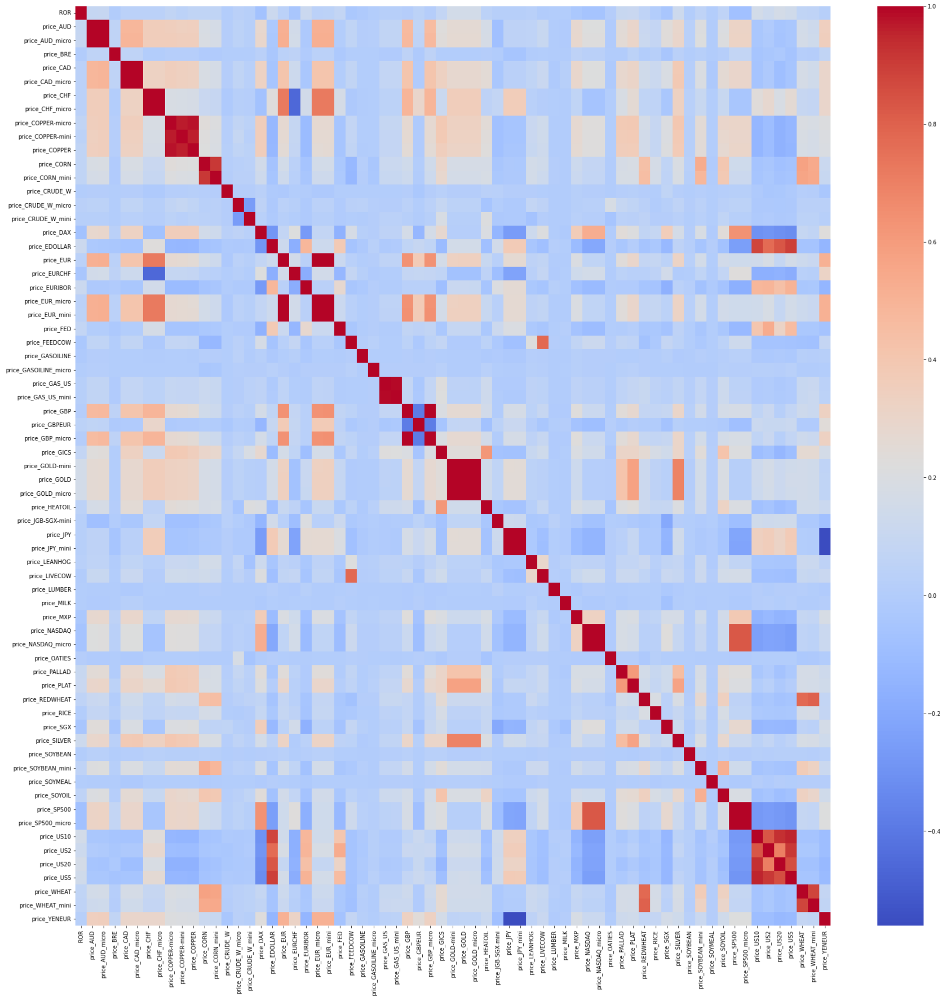

Figure 4: Heatmap.

Algorithm 1 Automated Instrument Selection

procedure SELECTINSTRUMENTS(N)
    Normalize the returns for volatility
    for each instrument i **do**
        Calculate the daily percentage returns
        Normalize returns by dividing by the standard deviation
    end for
    Randomly select the first instrument, i_1
    $selected \leftarrow \{i_1\}$
    for $j \leftarrow 2$ to N **do**
        $i_j \leftarrow \arg\min_{i \in \text{remaining instruments}} \text{AvgCorrelation}(\text{percentage returns of } i, selected)$
        $selected \leftarrow selected \cup \{i_j\}$
    end for
    return $selected$
end procedure

employs a rolling regression framework to effectively analyze and adjust the return ratios for SG CTA.

Rolling Regression involves a dynamic application of linear regression within a sliding time window. It entails fitting the regression model to a subset of the data within the window, testing it, and then moving the window forward and repeating the process. The goal is to capture temporal changes in relationships over time.

The rolling regression formula can be expressed as:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_p x_{p,t} + \varepsilon_t$$

Where y_t is the dependent variable at time t, $x_{1,t}, x_{2,t}, \ldots, x_{p,t}$ are the independent variables at time t, and $\beta_0, \beta_1, \ldots, \beta_p$ are the coefficients to be estimated.

The testing method involves the following steps:

1. Conduct a linear regression within the current time window to obtain a series of instrument weights.

2. Apply these weights to the instruments within the subsequent date of the same window, calculating the daily return for that date.

3. Slide the time window downwards and perform linear regression within the new window.

4. For the next date within the new window, repeat the procedure of calculating the daily return as in step 2.

5. Repeat the above steps iteratively.

6. After computing the daily return for the last date, return all of the daily, weekly and monthly returns.

Algorithm 2 provides a pseudocode representation of the rolling regression procedure:

Algorithm 2 Rolling Regression

```
procedure RollingRegression(x_values, y_values, window_size)
    for i ← window_size to length(x_values) do
        x_window ← x_values[i − window_size : i]
        y_window ← y_values[i − window_size : i]
        Perform linear regression on x_window and y_window to obtain weights
        current_date_x ← x_values[i]
        current_date_y ← y_values[i]
        Compute daily return using obtained weights and current_date_x
    end for
    return Daily, weekly, and monthly returns
end procedure
```

To assess the robustness of our replication method under various choices of these parameters, we conducted numerous experiments, each featuring different numbers of instruments and window sizes. A thorough statistical analysis followed these experiments, aimed at understanding how changes in these parameters affected the replication's performance. The key objective was to evaluate whether our replication approach remains effective and reliable across a range of different settings. This approach ensures that our findings are not just a result of overfitting to a particular set of parameters or instruments but are indicative of the general robustness and adaptability of the replication method.

3.4.1 Regularization methods

Considering the large number of variables (10-50 instruments), regularization techniques have been employed to mitigate the risk of overfit. We explored Lasso, Ridge, and Elastic Net regularization methods to enhance the robustness of our analysis.

Lasso, or Least Absolute Shrinkage and Selection Operator, is a regularization technique that encourages sparsity in the regression coefficients. Its mathematical formulation is as follows:

$$\text{Lasso:} \quad \min_{\beta} \left(\sum_{i=1}^{n} (y_i - \mathbf{x}_i \cdot \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

Here, λ is a tuning parameter that controls the strength of the penalty applied to the size of the coefficients. A higher value of λ increases the penalty, leading to more coefficients being shrunk towards zero Ridge regression, on the other hand, addresses multicollinearity by adding a penalty term proportional to the square of the coefficients:

$$\text{Ridge:} \quad \min_{\beta} \left(\sum_{i=1}^{n} (y_i - \mathbf{x}_i \cdot \beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right)$$

In Ridge regression, λ serves a similar purpose as in Lasso, but it affects the coefficients differently. Instead of shrinking coefficients completely to zero, it reduces their magnitude, thereby mitigating multicollinearity issues.

Elastic Net regression combines both Lasso and Ridge penalties, providing a balance between them:

$$\text{Elastic Net:} \quad \min_{\beta} \left(\sum_{i=1}^{n} (y_i - \mathbf{x}_i \cdot \beta)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right)$$

In Elastic Net, λ_1 and λ_2 are tuning parameters controlling the balance between Lasso and Ridge regularization, with λ_1 governing the Lasso component and λ_2 the Ridge component.

Through the visualization of the returns, we observed that Elastic Net is notably more effective than Ridge and Lasso. For example, Ridge appears to be sensitive to outliers, as depicted in Figure 5. ElasticNet combines the advantages of L1 and L2 regularization. By mixing the penalty terms of Lasso and Ridge, it can perform feature selection (through L1 regularization) and handle multicollinearity (through L2 regularization). This combination makes ElasticNet more stable and effective than using Lasso or Ridge alone, especially when dealing with high-dimensional data.

Figure 5: Effect of Ridge Regularization

3.5 Flow of the top-down method

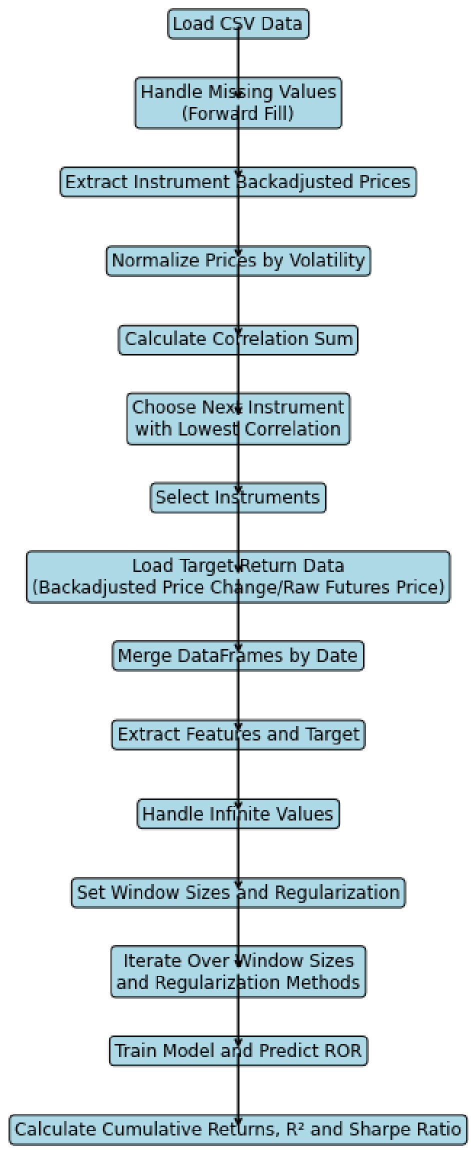

Figure 6: Flow of the top-down method

Figure 6 is the flowchart of the top-down method. We start by loading backadjusted prices, followed by handling missing values through forward filling. Next, we extract the instrument prices and normalize them by their volatility. After this, we calculate the average correlation and select the next instrument with the lowest correlation, repeating this until we have selected seven instruments. We then load the target return data and merge it with our selected instruments by date. The features and target variables are extracted, and infinite values are handled. We set the window sizes and regularization methods, iterating over these to train the model and predict the return. Subsequently, we calculate the monthly R^2 and Sharpe ratio. Finally, we compute the cumulative returns and plot them to visualize the model's performance over time.

4 The bottom-up method

Compared to the top-down approach, the bottom-up approach takes a different perspective when it comes to portfolio construction. Instead of treating the portfolio's components as a black box and attempting to infer their characteristics through reverse engineering, the bottom-up method relies on historical price data of individual instruments and employs statistical techniques to predict whether an instrument's price is likely to increase or decrease. The bottom-up approach has several key distinctions, particularly in its emphasis on risk management, especially in position sizing.

This section starts by detailing position sizing based on risk parity. It then introduces the trend-following method and its application. The discussion proceeds to the use of the carry strategy in futures trading. Finally, the section concludes with an explanation of how optimization techniques can fine-tune the forecast weights of both the trend-following and carry strategies, resulting in a composite approach that integrates elements of both.

4.1 Risk management

4.1.1 Measuring risk

In futures trading, leverage increases the risks. Because of this, the goals of CTAs are not just about making profits; they also include carefully measuring and managing risks. One way to measure this risk is by using standard deviation. The formula for simple standard deviation is as follows:

1. **Calculate the Mean:**

$$\bar{r} = \frac{\sum_{i=1}^{T} r_i}{T}$$

2. **Calculate Variance:**

$$\text{Variance} = \frac{\sum_{i=1}^{T} (r_i - \bar{r})^2}{T}$$

3. **Compute Standard Deviation:**

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$

Another approach to quantify risk is the use of exponentially weighted standard deviation. The exponentially weighted standard deviation is a measure that gives more weight to recent observations while calculating the variability of a dataset. It can be defined in terms of the span parameter.

Given a time series x_t, the exponentially weighted variance σ_t^2 at time t is defined as:

$$\sigma_t^2 = \frac{\sum_{i=0}^{t} w_i (x_i - \mu_t)^2}{\sum_{i=0}^{t} w_i} \tag{8}$$

where:

- $w_i = (1-\alpha)^{t-i}$ for $0 \leq i \leq t$
- $\alpha = \frac{2}{\text{span}+1}$
- $\mu_t = \frac{\sum_{i=0}^{t} w_i x_i}{\sum_{i=0}^{t} w_i}$ is the exponentially weighted mean at time t

To ensure the weights are normalized, we use the fact that the sum of weights should be 1 when $t \to \infty$:

$$\sum_{i=0}^{\infty} w_i = 1 \tag{9}$$

When $t = 0$, the exponentially weighted variance is not defined, so we initialize it as:

$$\sigma_0^2 = 0 \tag{10}$$

The exponentially weighted standard deviation σ_t is the square root of the exponentially weighted variance:

$$\sigma_t = \sqrt{\sigma_t^2} \tag{11}$$

The exponentially weighted standard deviation at time t can be summarized as:

$$\sigma_t = \sqrt{\frac{\sum_{i=0}^{t}(1-\alpha)^{t-i}(x_i - \mu_t)^2}{\sum_{i=0}^{t}(1-\alpha)^{t-i}}} \tag{12}$$

where $\alpha = \frac{2}{\text{span}+1}$ and μ_t is the exponentially weighted mean.

Another method involves combining the standard deviations of long-term and short-term data to improve forecasting. This approach reduces the risk of large position changes just before a crisis and lowers trading costs due to the stabilizing effect of the long-term estimate. The long-term estimate smooths out volatility, providing protection against sudden spikes. This method works because while short-term volatility might be low, long-term data shows potential maximum volatility. By averaging long-term and short-term volatilities, this blended measure prevents underestimating risk when recent volatility is unusually low, ensuring more cautious position sizing.

The blended exponentially weighted moving average of standard deviations is formulated as:

$$\sigma = w_{\text{short}} \times \sigma(\text{span}_{\text{short}}) + w_{\text{long}} \times \sigma(\text{span}_{\text{long}}) \tag{13}$$

Where:

- σ represents the blended standard deviation.
- $\sigma(\text{span}_{\text{short}})$ is the exponentially weighted standard deviation with a short-term window $\text{span}_{\text{short}}$ (e.g., 32 days).
- $\sigma(\text{span}_{\text{long}})$ is the exponentially weighted standard deviation with a long-term window $\text{span}_{\text{long}}$ (e.g., 1000 days).
- w_{short} and w_{long} are the weights applied to the short-term and long-term standard deviations, respectively.

In practice, the weights w_{short} and w_{long} are chosen such that $w_{\text{short}} + w_{\text{long}} = 1$ and that w_{short} and $w_{\text{long}} > 0$. For instance, one might set $w_{\text{short}} = 0.7$ and $w_{\text{long}} = 0.3$ to reflect a higher emphasis on recent market conditions while still accounting for long-term volatility patterns.

4.1.2 Position scaling

Risk management is a cornerstone of our trading strategy. We apply principles of risk parity for scaling positions, where each position is sized according to its contribution to the overall portfolio risk. The objective is to optimize risk distribution across the portfolio. The target risk is set as a proportion of the portfolio's total value, and in our analysis, we estimate this level based on historical data. For reference, the SG CTA index has a historical volatility of around 10%, which we use as a guide for setting our target risk.

4.2 Trend

The trend-following method is utilized to determine the direction of futures prices and to inform trading decisions. Historical data indicate that many markets exhibit momentum within approximately one-week to one-year time horizons. The aim of a trend-following signal is to capitalize on this momentum behavior. By generating trading signals based on this principle, the strategy seeks to profit from the continuation of price trends.

The presence of momentum is detected by calculating the EWMA over a chosen period. A current price that is higher than the EWMA suggests an uptrend and typically triggers a buy signal. However, a drawback of this approach is its susceptibility to frequent fluctuations in price movements, leading to excessively frequent trading signals. To address this issue, the exponentially weighted moving average crossover (EWMAC) is computed, and trading decisions are based on its positive or negative values. EWMAC is derived by subtracting the long-term EWMA from the short-term EWMA. For instance, in the calculation of a crossover with a short-term duration of 64 days and a long-term duration of 256 days, the formula is as follows:

$$\text{EWMAC}(64, 256) = \text{EWMA}(span = 64) - \text{EWMA}(span = 256).$$

If EWMAC is greater than 0, a buying signal is generated, whereas if it is less than 0, a selling signal is indicated. This approach mitigates the issue of excessive trading signals associated with abrupt price fluctuations.

As what we discussed in previous section, EWMA can be expressed as:

$$\text{EWMA}_t = \frac{2}{\text{span} + 1} \cdot x_t + \left(1 - \frac{2}{\text{span} + 1}\right) \cdot \text{EWMA}_{t-1} \tag{14}$$

with the initial condition:

$$\text{EWMA}_0 = x_0 \tag{15}$$

4.2.1 Trend forecast

Using EWMAC as a method to determine whether prices are in an uptrend is indeed effective. However, considering the variations in prices among different instruments, EWMAC values are likely to differ significantly, making it challenging to compare trend signals across diverse instruments and to combine them coherently. To address this issue, we introduce a slight modification to EWMAC - dividing the crossover by the standard deviation of returns, which represents the risk calculated in units of daily price. This adjustment results in the creation of a risk-adjusted EWMAC, referred to as the "raw forecast" in subsequent subsections. The formula for this signal-to-noise ratio is as follows:

$$\text{Raw forecast} = \frac{\text{Fast EWMA} - \text{Slow EWMA}}{\sigma_p}$$

where

$$\sigma_p = \text{Price} \times \frac{\sigma\%}{16}$$

and $\sigma\%$ represents the annualized percentage risk.

The risk in daily price points is equal to the price multiplied by the annualised percentage risk, divided by 16. This refined approach allows for a more accurate assessment of uptrends while accounting for the inherent risk associated with daily price fluctuations.

The forecast's sign clearly determines our trading position: a positive forecast suggests a long position, while a negative forecast implies a short position. A forecast above the average size corresponds to a larger position, whereas a forecast below the average results.

To get a more accurate estimate for the average value, we can measure it across many different instruments and take an average of those estimates. We standardize our forecasts to have an average absolute value of 10. This means that after determining the average value of the raw forecast, we scale each individual forecast accordingly. The scaled forecast is calculated by multiplying the raw forecast by 10 and then dividing it by the average absolute value of the raw forecast. This scaling process ensures that our forecasts are consistently measured and comparable, facilitating a more accurate assessment of market trends and guiding our trading decisions effectively. The scaled forecast is calculated using the formula:

$$\text{Scaled forecast} = \frac{\text{Raw forecast} \times 10}{\text{Average absolute value of raw forecast}}$$

In order to avoid extreme values, the forecast is capped using the following formula:

$$\text{Capped forecast} = \text{Max}\left(\text{Min}\left(\text{Scaled forecast}, +20\right), -20\right)$$

Where the Scaled forecast represents the initial forecast value. This method ensures that the forecasted values are constrained within a reasonable range, preventing excessively large or small predictions.

We introduce the notations for various trend forecasts using different spans for EWMAs:

- Trend2: Fast EWMA with span of 2, Slow EWMA with span of $2 \times 4 = 8$.
- Trend4: Fast EWMA with span of 4, Slow EWMA with span of $4 \times 4 = 16$.
- Trend8: Fast EWMA with span of 8, Slow EWMA with span of $8 \times 4 = 32$.
- Trend16: Fast EWMA with span of 16, Slow EWMA with span of $16 \times 4 = 64$.
- Trend32: Fast EWMA with span of 32, Slow EWMA with span of $32 \times 4 = 128$.
- Trend64: Fast EWMA with span of 64, Slow EWMA with span of $64 \times 4 = 256$.

Furthermore, the optimal position can be calculated using the following formula:

$$N_{i,t} = \text{Capped forecast} \times \text{Capital} \times \text{Weight}_i \times \tau \div (\text{Multiplier}_i \times \text{Price}_{i,t} \times \text{FX rate}_{i,t} \times \sigma\%_{i,t})$$

Here, $N_{i,t}$ represents the optimal number of contracts, Capped forecast is the capped scaled forecast of trend signals, Capital denotes the available capital, Weight_i represents a specific weight associated with instrument i, τ signifies the target risk, Multiplier_i represents a multiplier associated with instrument i, $\text{Price}_{i,t}$ represents the price of instrument i at time t, FX $\text{rate}_{i,t}$ is the foreign exchange rate at time t, and $\sigma\%_{i,t}$ represents the annualized percentage risk associated with instrument i at time t.

4.2.2 Flow of Trend

Figure 7 is the flowchart of Trend. We start by setting input and output paths, where input folders contain trend and carry price data, and the output folder is for saving results. We iterate over each CSV file of backadjusted prices in US dollar, reading and preprocessing the data by filling missing values and formatting dates. Adjusted and current price data are then merged by date. We calculate daily price changes and percentage changes to determine daily

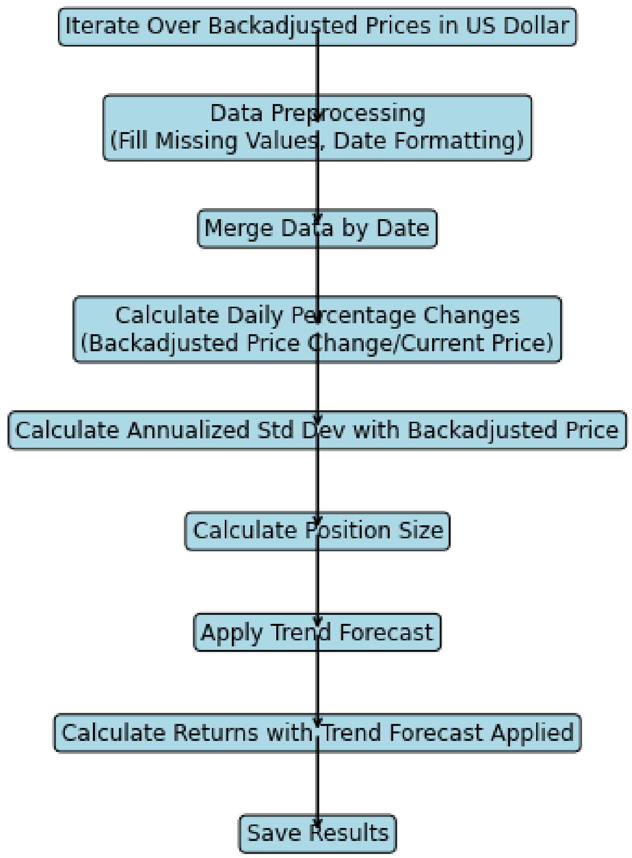

Figure 7: Flow of Trend

returns. The algorithm computes the exponential moving standard deviation of these returns and annualizes it to understand annual volatility. Using this annualized standard deviation, capital, and risk target, we determine the position size. A trend forecast is applied using fast and slow EWMAs, adjusted by the standard deviation for a risk-adjusted trend forecast. We calculate returns by applying position size and the trend forecast on the adjusted prices. Finally, we save these percentage returns into new CSV files in the output folder, effectively implementing the Trend Strategy.

4.3 Carry

4.3.1 Measuring Carry

The raw carry is determined by the price difference between the nearer futures contract and the currently held contract. Carver [2023a] gives the formula below:

$$\text{Raw carry} = \text{Price of nearer futures contract} - \text{Price of currently held contract}$$

However, when the currently held contract is close to expiration and there is no nearer futures contract available, the raw carry is computed as the difference between the prices of the currently held contract and the further out contract. Carver [2023a] gives the formula to calculate raw carry.

$$\text{Raw carry} = \text{Price of currently held contract} - \text{Price of further out contract}$$

To standardize raw carry values for comparison, we annualize the raw carry by considering the difference in expiration time between contracts in years. Carver [2023a] gives the formula to annualize the raw carry.

$$\text{Expiry difference in years} = \frac{\text{abs(months between contracts)}}{12}$$

The annualized raw carry is then calculated as:

$$\text{Annualized raw carry} = \frac{\text{Raw carry}}{\text{Expiry difference in years}}$$

For risk management purposes, we convert the annualized raw carry into a risk-adjusted form by dividing it by the annualized standard deviation of returns for the relevant instrument. According to Carver [2023a], the risk-adjusted carry (Carry) is given by the formula:

$$\text{Carry} = \frac{\text{Annualized raw carry}}{\sigma\% \times \text{Current contract price}}$$

Here, $\sigma\%$ represents the estimated annual standard deviation of returns expressed in percentage terms.

4.3.2 Carry forecast

To implement carry into trading strategies, we introduce a carry forecast, a signal to noise ratio that is proportional to an expected risk-adjusted return. According to Carver [2023a], the Raw Carry Forecast is expressed as:

$$\text{Carry Raw Forecast} = \frac{\text{Annualized Raw Carry}}{\sigma_p \times 16}$$

Here, σ_p denotes the daily standard deviation of price differences.

However, the presence of noise in forecasts can lead to unneccessary trades. To mitigate this issue, a smoothing technique is employed, employing an EWMA. The smoothed Carry Forecast with a specified smoothing parameter, known as the Span, is calculated using the following formula give by Carver [2023a]:

$$\text{Smoothed Carry Forecast(Span)} = \text{EWMA}_{\text{Span}}(\text{Carry Forecast})$$

To facilitate the comparison of different instruments and time spans in carry analysis, we standardize the scaled carry, akin to a trend strategy, to a comparable magnitude. The formula is as follows:

$$\text{Scaled carry forecast} = \text{Smoothed carry forecast} \times \text{Forecast scalar}$$

The Forecast scalar used here for scaling the carry forecast aligns with the one utilized in trend analysis, ensuring consistency between trend and carry signals.

In order to prevent extreme values, the scaled carry forecast is capped using the following operation:

$$\text{Capped carry forecast} = \text{Max}\left(\text{Min}\left(\text{Scaled carry forecast}, +20\right), -20\right)$$

This capping mechanism confines the carry forecast within reasonable boundaries, averting excessively large or small predictions. We introduce the notations for various carry forecasts using different spans for the EWMAs:

- Carry5: Smoothed carry forecast with span of 5.
- Carry20: Smoothed carry forecast with span of 20.
- Carry60: Smoothed carry forecast with span of 60.
- Carry120: Smoothed carry forecast with span of 120.

Subsequently, the capped carry forecast is employed to compute the optimal position size:

$$N_{i,t} = \text{Capped carry forecast} \times \text{Capital} \times \text{Weight}_i \times \tau \div (\text{Price}_{i,t} \times \text{FX rate}_{i,t} \times \sigma\%_{i,t})$$

Here, $N_{i,t}$ represents the optimal number of contracts, Capped carry forecast is the capped and scaled carry forecast, Capital denotes the available capital, Weight_i represents a specific weight associated with instrument i, τ signifies the target risk, $\text{Price}_{i,t}$ represents the price of instrument i at time t, FX $\text{rate}_{i,t}$ is the foreign exchange rate at time t, and $\sigma\%_{i,t}$ represents the annualized percentage risk associated with instrument i at time t.

4.3.3 Flow of Carry

Figure 8 is the flowchart of Carry. First, we set the input and output folder paths. The input folder contains trend and carry price data, and the output folder is used to save the results. Then, we iterate over each CSV file in the input folder, reading the adjusted prices and current prices data. Next, we preprocess the data by filling missing values (using forward fill) and formatting the dates. After that, we merge the adjusted prices data and current prices data by date to generate a dataframe with all necessary information. Following this, we calculate daily price changes and daily percentage changes, as well as the exponential moving standard deviation of daily returns, which is then annualized. We calculate the ten-year volatility and combine it with the annualized standard deviation to obtain a stabilized volatility measure.

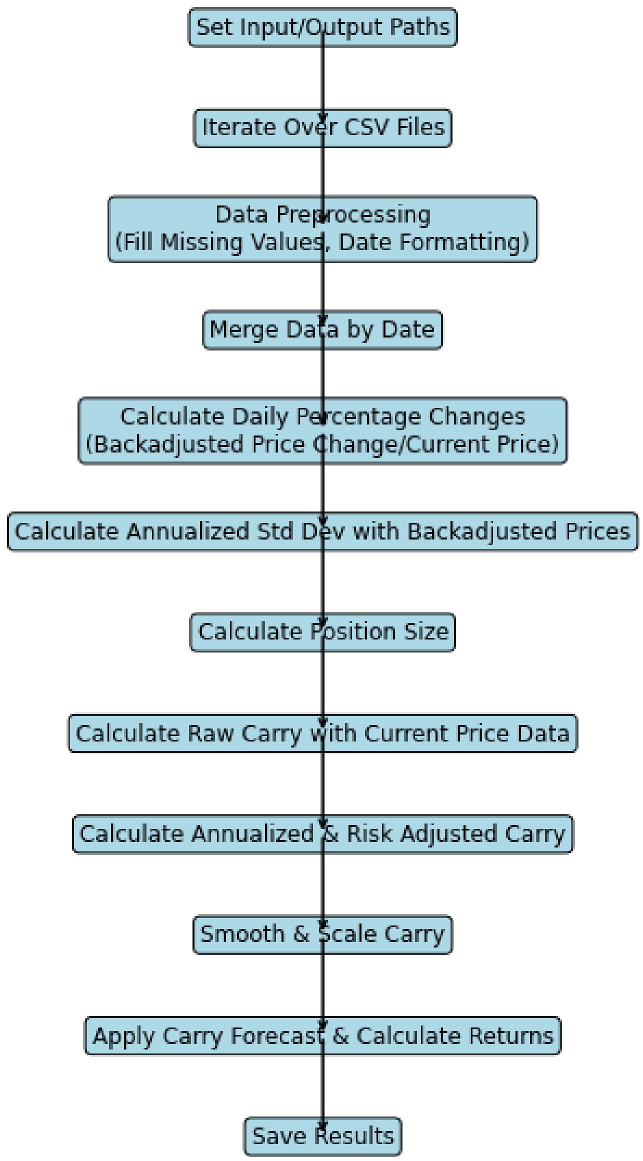

Figure 8: Flow of Carry

Based on the volatility-adjusted risk target, we calculate the position size, and then compute the raw carry, annualized carry, and price volatility. Next, we calculate the risk-adjusted carry. We use exponential weighted moving average to smooth the carry, then scale it and cap it within a specified range. Subsequently, we apply the carry forecast to calculate the positions and daily price change returns. Finally, we save the calculated percentage returns to new CSV files.

4.4 Optimization

In order to explore the weights of various instruments, as well as the weights of different carry or trend signals, we use optimization techniques to address this challenge. This approach is feasible due to the linearity of the impact of instruments and carry or trend signals on position scaling as well as returns. Moreover, the forecasts we construct are all scaled to the same order of magnitude, facilitating their integration.

4.4.1 Objective functions and constraints

Our objective is to minimize the tracking error between SG CTA's index and the overall returns derived from our bottom-up method:

Objective Function:

$$\text{Tracking Error} = |\text{SG CTA returns} - \text{Overall returns}|$$

$$\text{Overall returns} = \sum_{i \in \text{instruments}} v_{\text{asset-class}(i)} \sum_{s \in \text{signals}} \tau \cdot w_s \times \text{returns}(s) \tag{16}$$

where:

$$\begin{aligned} \tau &\text{ is the risk target parameter} \\ v_{\text{asset-class}(i)} &\text{ is the weight for each asset class of instrument } i \\ w_s &\text{ is the weight of the signal } s \\ \text{returns}(s) &\text{ is the return of the signal } s \end{aligned}$$

$$\text{instruments} = \{\text{instrument}_1, \text{instrument}_2, \ldots, \text{instrument}_n\} \tag{17}$$

$$\begin{aligned} \text{signals} = \{&\text{Trend2}, \text{Trend4}, \text{Trend8}, \text{Trend16}, \text{Trend32}, \text{Trend64}, \\ &\text{Carry5}, \text{Carry20}, \text{Carry60}, \text{Carry120}\} \end{aligned} \tag{18}$$

Constraints:

$$\begin{aligned} &w_s > 0 \quad \text{(Weights of trend and carry signals are greater than zero)} \\ &\sum_{s \text{ in signal}} w_s = 1 \quad \text{(Sum of weights of signals equals 1)} \\ &v_a > 0 \quad \text{for all asset classes } a \\ &\sum_{a \in \text{asset-classes}} v_a = 1 \end{aligned}$$

Instruments refer to the specific financial assets included in our portfolio. These instruments are categorized into different asset classes, such as agricultural products, bonds, commodities, energy, metals, and currencies. Each asset class contains multiple instruments, and we assume a constant weight for each instrument within the same asset class. We jointly optimize the weights for the asset classes and the weights for the signals. This forms a non-linear optimization problem, and we use a general-purpose gradient-based optimizer to solve it.

4.4.2 Flow of optimization

Figure 9 is the flowchart of optimization. We begin by reading the CSV files containing the data, followed by filling any NaN and infinite values to ensure data integrity. Next, we set the date index and filter the data to include only relevant records. We then extract the features and target variables, checking again for any NaN or infinite values. Parameters are randomly initialized to prepare for the optimization process. The training dataset is selected, and the data is converted to PyTorch tensors for compatibility with machine learning models. We initialize the parameters in log space to maintain positive values and define the objective function. An Adam optimizer is initialized, and minibatches are created for efficient training. The core of the process involves an optimization loop with minibatches and gradient clipping to refine the parameters iteratively. Once optimized, the weights are saved to a CSV file. Finally, we calculate the R^2 and Sharpe ratio to evaluate the model's performance.

4.4.3 Overfitting

According to Hawkins [2004], a minimum of 10 observations per term in a linear model is essential for reliable results. Our study's training set includes 6144 observations, which, with a maximum of 50 terms, equates to roughly 122.8 observations per term. Even when replicating with a subset of 10 years, which has the least training data, there are about 40 observations per term. These figures, well above the recommended threshold of 10, indicate that our study is not prone to overfitting.

4.5 Conclusions

This subsection provides an overview of how the bottom-up approach is employed to replicate CTAs, with a primary focus on trend and carry strategies. The utilization of optimization techniques allows us to determine the optimal instrument weights and signal weights. Subsequently, the next subsection will present the results and conduct a comparative analysis between the bottom-up and top-down approaches.

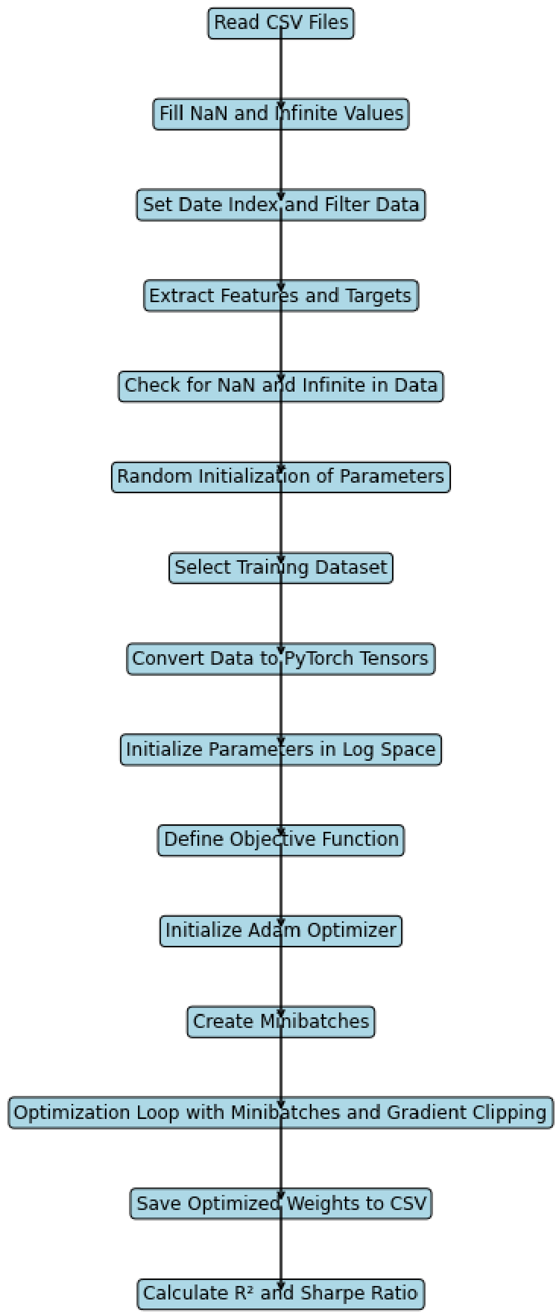

Figure 9: Flow of optimization

5 Results and analysis

The previous sections delved into the methods and details of the top-down and bottom-up approaches. This section now aims to showcase and compare the results of these approaches, followed by a comprehensive analysis.

5.1 Results of top-down methods

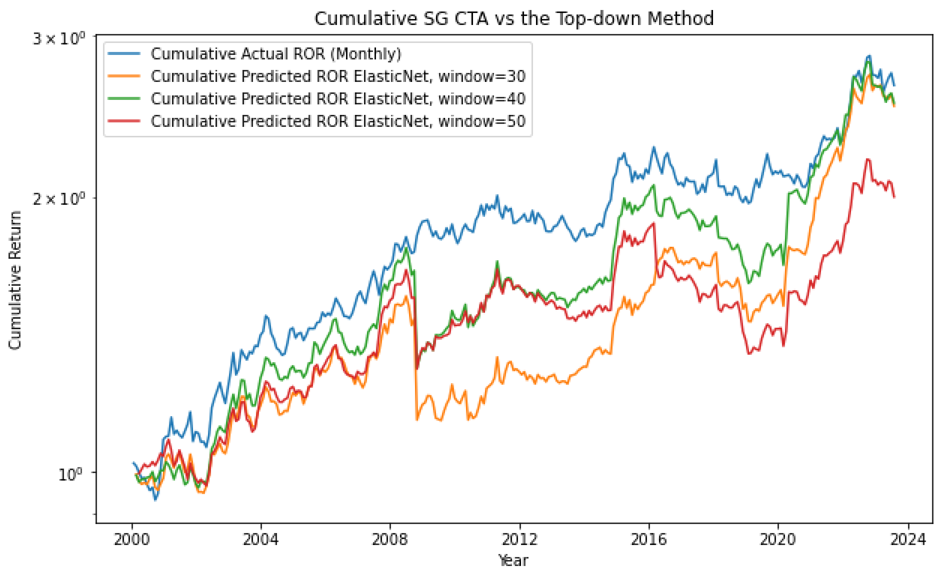

Figure 10: Cumulative Returns

Figure 10 shows the results of replicating the cumulative returns of SG CTA using the top-down method by randomly selecting 28 instruments. The graph compares the cumulative returns predicted by the ElasticNet model with different rolling time windows to the actual cumulative returns. We also tested the Lasso and Ridge methods, and found that ElasticNet performed the best. The Sharpe ratio for ElasticNet 30, ElasticNet 40, and ElasticNet 50 are 0.47, 0.48, and 0.39. The green curve has the best fit. The green curve represents the cumulative returns predicted by the ElasticNet model with a 30-day rolling time window. A 30-day rolling window means that the model uses the daily returns from the past 30 days to train the regression model and predict the return for the next day. The ElasticNet regularization method combines the advantages of both Lasso and Ridge regularizations, aiming to improve the model's predictive power and stability. This curve closely matches the actual cumulative returns (blue curve) for some of the period, especially before 2008 and after 2020, showing relatively stable performance with moderate fluctuations.

To explore how window sizes and the number of instruments influence the replication, we performed multiple replications with the window size ranging from 10 to 40, and the number of instruments ranging from 20 to 50.

Figure 11 shows the relationship between window size (X-axis) and two metrics: the average R^2 (top plot) and the average Sharpe ratio (bottom plot). The number of instruments ranges from 20 to 50. The R^2 value is a measure of the goodness of fit of a regression model, typically ranging from 0 to 1, with higher values indicating better model fit. In the top plot, as the window size increases from 20 to 50, the average R^2 generally fluctuates. It starts at around 0.2 for window sizes 20 to 25, then peaks at approximately 0.37 when the window size is between 30 and 35. After a drop at window size 35, it rises again, reaching another peak around 0.38 at window size 40, and finally stabilizes around 0.3 to 0.35 for window sizes 45 to 50.

In the bottom plot, the average Sharpe ratio initially decreases slightly from 0.32 at window size 20 to 0.25 at window size 25. It then increases sharply, peaking at around 0.47 for window sizes 30 to 35. After peaking, the Sharpe ratio fluctuates and decreases gradually, stabilizing around 0.3 for window sizes 45 to 50.

Overall, the risk-adjusted returns of the portfolio are best when the window size is between 30 and 35, with the highest Sharpe ratio observed in this range. The R^2 value also indicates significant variation, peaking around the same window sizes.

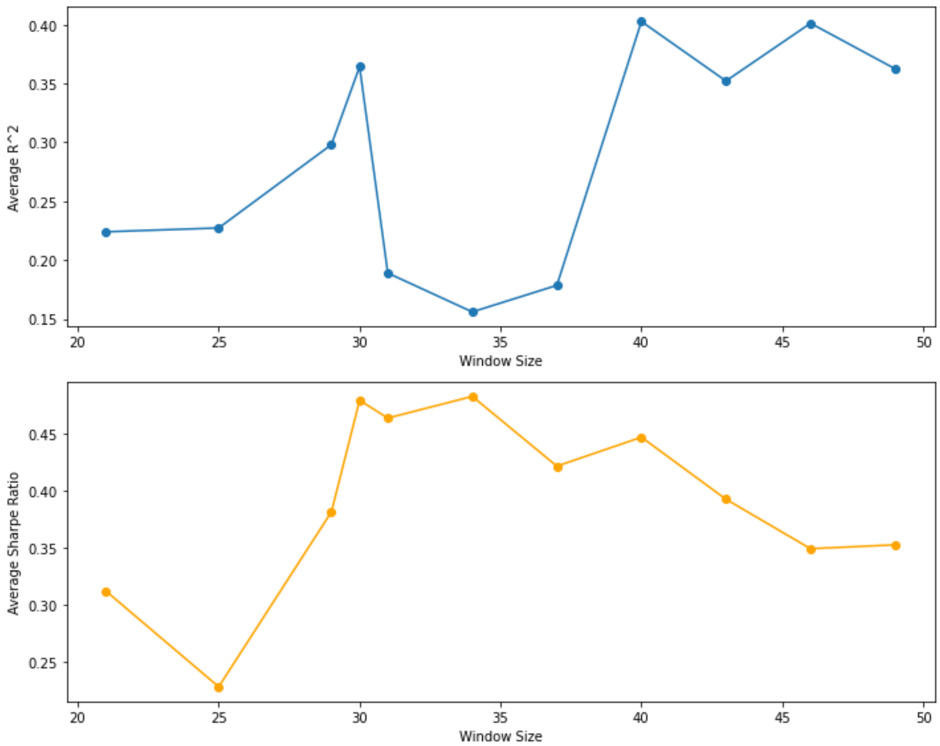

Figure 11: R^2 and Sharpe ratio by window size

Figure 12 shows the relationship between the number of instruments (X-axis) and two metrics: the average R^2 (top plot) and the average Sharpe ratio (bottom plot). The window size ranges from 10 to 40.

In the top plot, as the number of instruments increases from 10 to 35, the average R^2 initially decreases from around 0.35 at 10 instruments to approximately 0.15 at 20 instruments. It then starts increasing, reaching a peak at around 0.4 at 35 instruments.

In the bottom plot, the average Sharpe ratio also shows a fluctuating pattern. It starts at 0.25 for 10 instruments, dips to about 0.25 for 15 instruments, then increases, peaking at around 0.43 for 30 instruments. After the peak, it decreases to approximately 0.3 at 35 instruments.

Overall, the risk-adjusted returns of the portfolio are highest when the number of instruments is around 30, with the highest Sharpe ratio observed in this range. The R^2 value indicates a significant increase when the number of instruments approaches 35.

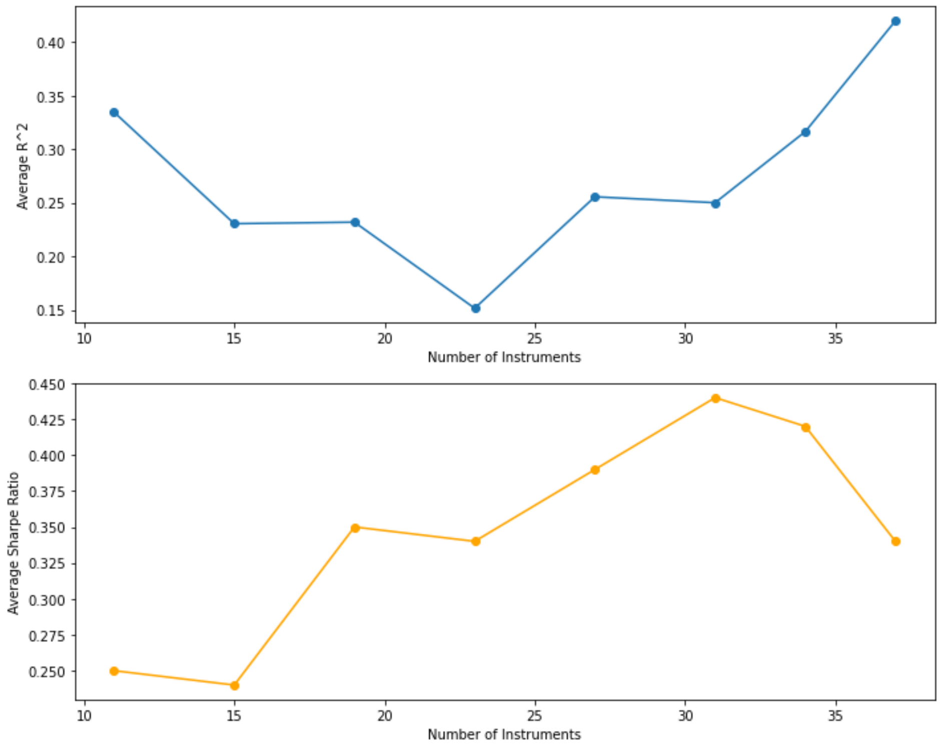

Figure 12: R^2 and Sharpe ratio by number of instruments

5.2 Robustness of the top-down method

To assess the robustness of the top-down method, we conducted cumulative return replications using the data of the mutual funds, including Capital [2024a], Beacon [2024], Fund [2024], AlphaSimplex [2024], Capital [2024b], Fidelity [2024], and PIMCO [2024]. Figures 13, 15, 16, 18, 17, 19, and 14 present the cumulative return replication results achieved through the top-down method for each of these CTA funds.

Figure 13 shows the results of replicating ABYIX using the top-down method, comparing

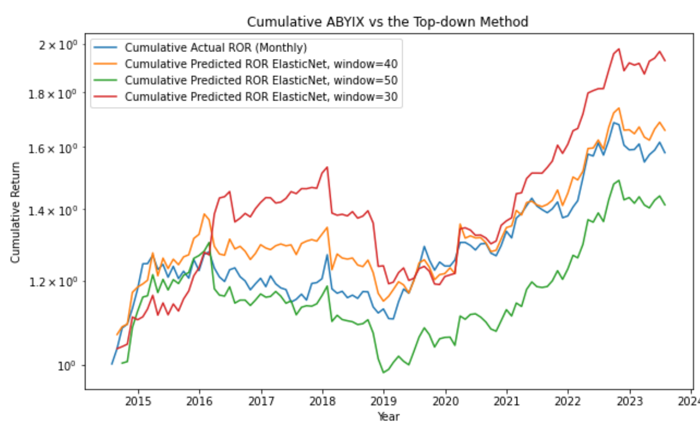

Figure 13: Cumulative returns of ABYIX and the top-down method

the cumulative actual returns (blue curve) with the predicted returns for different window sizes (orange, green, and red curves). Before 2019, the green curve (50-day window) had the best prediction performance, with the cumulative returns closely matching the actual returns, indicating that the 50-day window model performed the best during this period. After 2019, the orange curve (40-day window) had the best prediction performance, with the cumulative returns almost identical to the actual returns, indicating that the 40-day window model performed the best during this period.

Figure 14 illustrates the results of replicating PQTIX using the top-down method. From

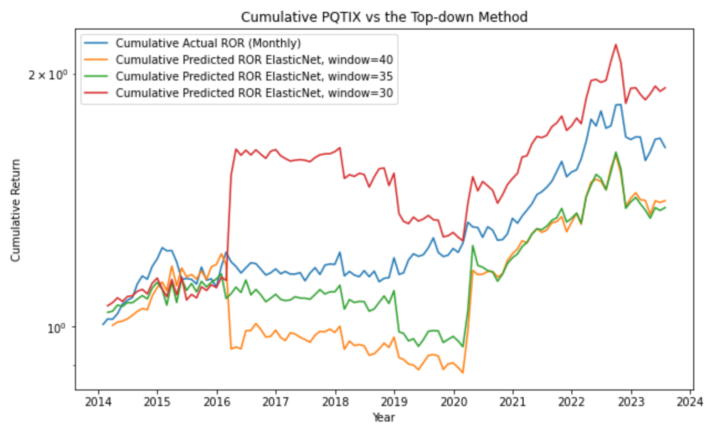

Figure 14: Cumulative returns of PQTIX and the top-down method

2014 to 2016, the actual returns and predicted returns were generally consistent, but there was a sudden change in 2016, leading to a significant decline in the model's predictive performance. During the period from 2019 to early 2020, the predicted returns again diverged from the actual returns, especially in early 2020. None of the models with different window sizes accurately predicted the market changes during this period, resulting in a significant increase in the gap between predicted and actual returns.

Figure 15 displays the results of replicating AHLYX using the top-down method. From 2015

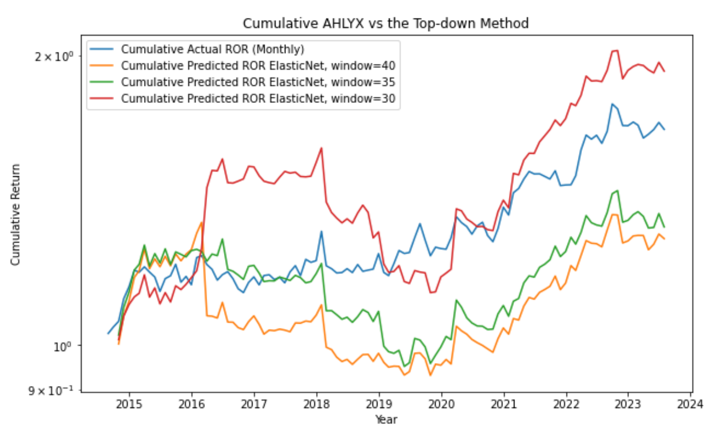

Figure 15: Cumulative returns of AHLYX and the top-down method

to 2016, the model's prediction results were mediocre, with a small gap between actual and predicted returns. In 2016, the gap between predicted and actual returns suddenly increased, indicating that the model failed to accurately capture the market's sharp changes. It was not until after 2019 that the models with 35-day and 40-day windows began to align more closely with actual returns, showing better predictive performance.

Figure 16 shows the results of replicating AQMIX using the top-down method. The green line (35-day window) model performs the best, capturing most of the market's changing trends. However, even though the green line performs well, there is always a certain degree of dif-

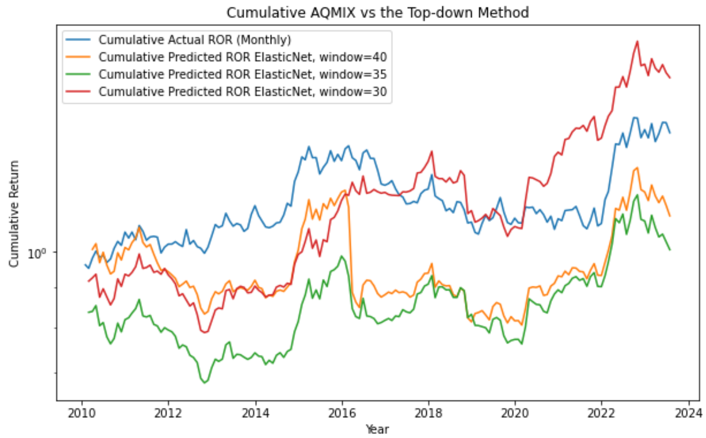

Figure 16: Cumulative returns of AQMIX and the top-down method

ference between it and the actual cumulative returns, failing to match the actual returns precisely.

Figure 17 presents the results of replicating EQCHX using the top-down method. The

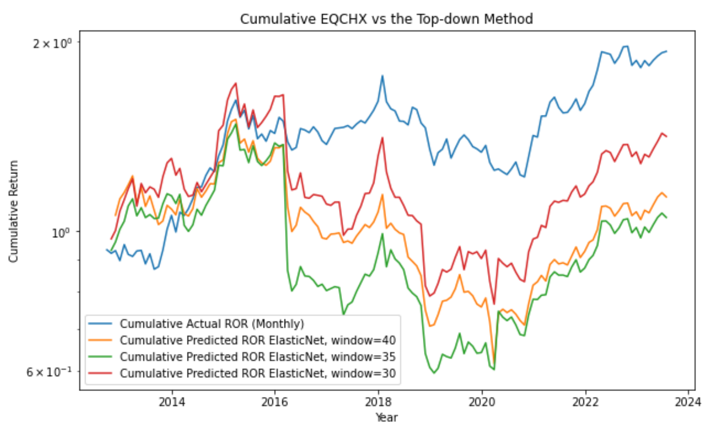

Figure 17: Cumulative returns of EQCHX and the top-down method

model's fit varies across different time periods. Before 2016, the model's prediction results were mediocre, with a small but not very accurate gap between actual and predicted returns. However, in 2016, the predicted returns suddenly changed, leading to a significant increase in the gap between the model's predictions and the actual returns. Since 2016, the model's prediction accuracy has improved.

Figure 18 shows the results of replicating ASFYX using the top-down method. From 2012 to 2013, the model's predictions closely matched the actual returns, indicating good performance. However, in 2013, there was a sudden shift in the predicted returns, resulting in a significant divergence from the actual returns. Another sharp change occurred in 2016, where the model again failed to accurately capture the market volatility. Despite these abrupt changes in 2013 and 2016, the model's predictions were generally aligned with the actual returns during other periods, showing relatively good performance.

Figure 19 illustrates the results of replicating MFTFX using the top-down method. Apart from the significant differences in 2012 and 2018, the model's predictions were generally consistent with the actual returns during the other periods. Particularly from 2020 onwards,

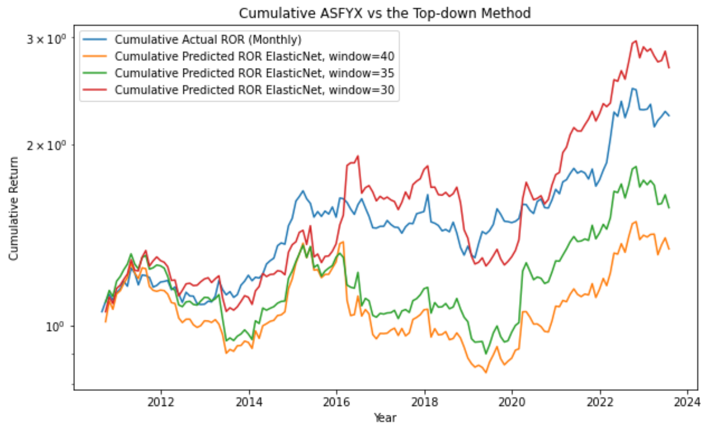

Figure 18: Cumulative returns of ASFYX and the top-down method

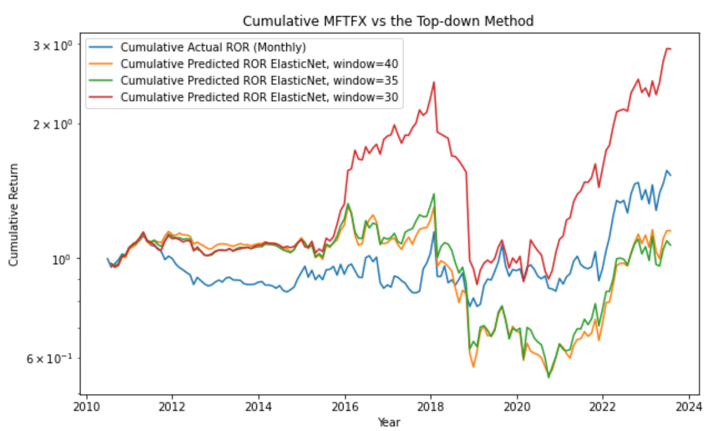

Figure 19: Cumulative returns of MFTFX and the top-down method

the 40-day window model performed well, accurately capturing market trends with minimal differences from the actual returns.

5.3 Results of bottom-up methods

In the bottom-up method, first we use trend following and carry strategies to calculate the return of each instruments. Then we optimize the asset class weights and signal weights for trend and carry strategies. We classify all of the instruments into 6 classes and take the average of the instruments' return under each asset class to be the asset class return. The 6 asset classes are: equities, bonds, currencies, energy, metals, and agricultural products.The classification of these instruments can be found in Table 1. Table 2 shows the optimal weights for trend and carry signals, while Figure 20 represents the optimal weights for each asset class. Furthermore, Figure 21 illustrates the cumulative returns of SG CTA and the cumulative returns achieved through the bottom-up method.

Table 1: Classification of instruments

Category	**Instruments**
Equities	AEX, CAC, DAX, DOW_mini, EU-BANKS, EU-DIV30, EURO600, FTSE100, FT-SECHINAA, MSCIEM-LIFFE, MSCISING, NASDAQ_micro, NIFTY, NIKKEI, SP400, SP500_micro, TOPIX, US-DISCRETE, US-FINANCE, US-HEALTH, US-MATERIAL, US-STAPLES, US-TECH, US-UTILS
Bonds	AUSCASH, BOBL, BTP, BTP3, BUND, BUXL, US10, US2, US20, US30, US5, GILT, JGB, SHATZ, EUROBOR-ICE, FED, CADSTIR, CAD10, CAD2, CAD5
Currencies	AUD_micro, CAD, CHF, DX, EUR_micro, GBP, GBPEUR, JPY, MXP, NZD, YENEUR, ZAR, BRE
Metals	AULIMINIUM_LME, COPPER, GOLD_micro, NICKEL_LME, PALLAD, PLAT, SILVER
Energies	BRENT-LAST, CRUDE_W_mini, GASOIL, GASOILINE, GAS_US_mini, HEATOIL
Agricultural Products	CANOLA, COCOA, COCOA_LDN, COFFEE, CORN_mini, COTTON2, FEED-COW, LEANHOG, LIVECOW, MILLWHEAT, RAPESEED, REDWHEAT, ROBUSTA, RUBBER, SUGAR11, SUGAR_WHITE, WHEAT_mini, SOY-BEAN_mini, SOYMEAL, SOYOIL

Signal	Weight
Trend2	0.0110
Trend4	0.3430
Trend8	0.0003
Trend16	0.0053
Trend32	0.1097
Trend64	0.0017
Carry5	0.0333
Carry20	0.0599
Carry60	0.3475
Carry120	0.0879

Table 2: Weights of signals

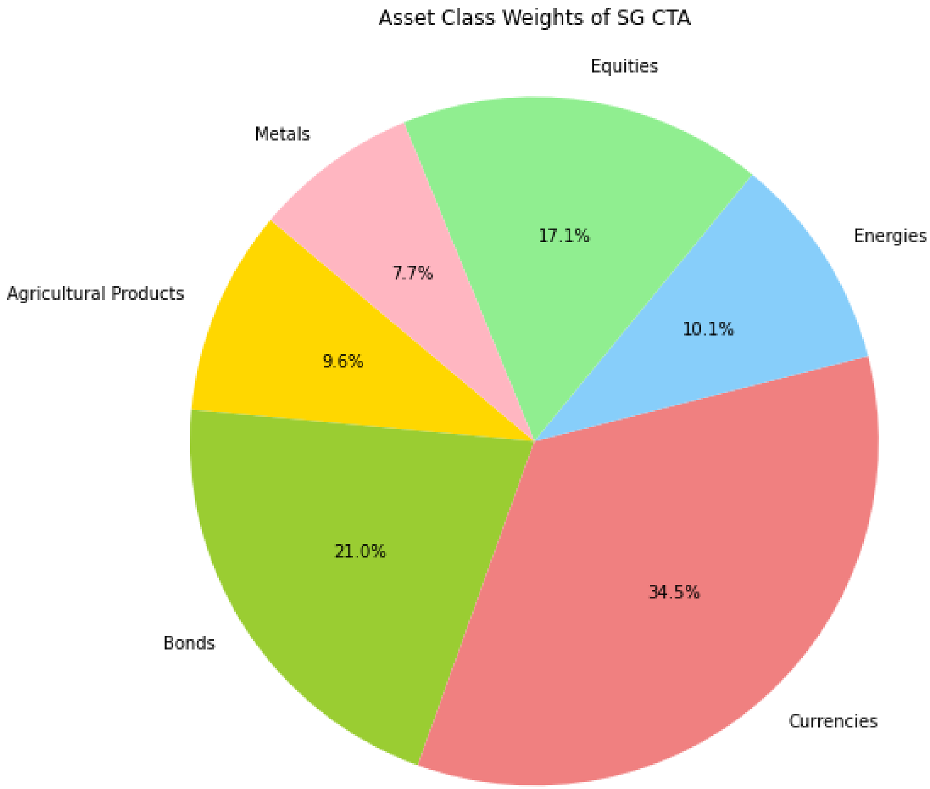

Figure 20: Asset allocation

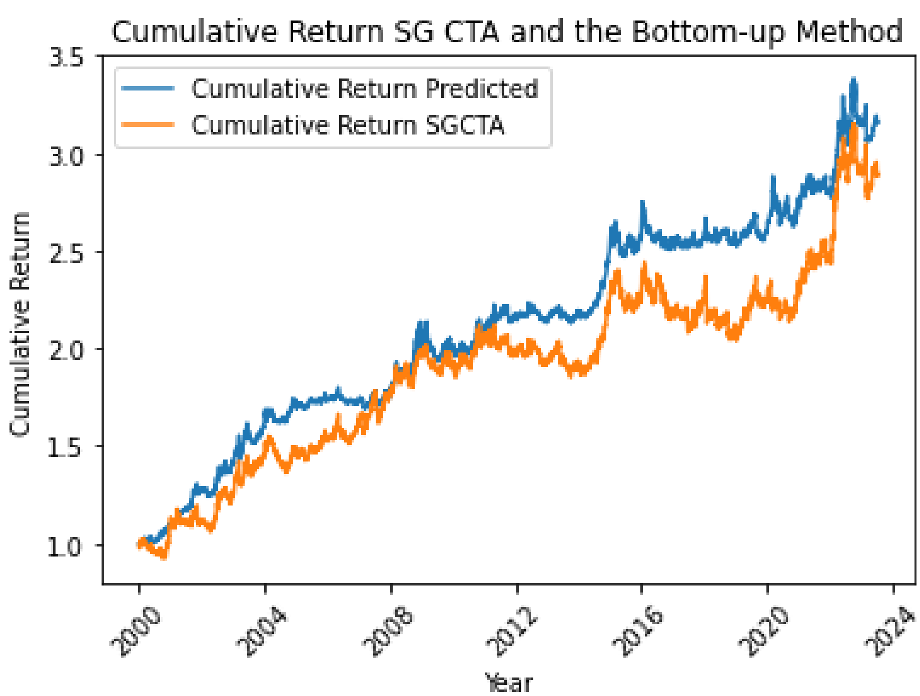

Figure 21: Cumulative Returns

With the optimal weights, the results show that the corresponding values for the Sharpe ratio and R^2 are 0.80 and 0.59. Figure 21 provides an overview of the cumulative return over time, offering a comprehensive view of the overall performance achieved through the bottom-up method. Overall, the model's fit varies across different time periods. From 2000 to 2004 and after 2016, the model's predictive performance is better, with cumulative predicted returns closely matching the actual returns. However, during 2005 to 2010 and 2011 to 2015, the model's predictive performance is poorer, showing significant deviations. The $\mathbf{R^2}$ value is 0.59, indicating that the model can explain approximately 59% of the variance in actual returns. The Sharpe ratio is 0.80, indicating that the model has good risk-adjusted returns.

5.4 Robustness of the bottom-up method

To check the robustness of the bottom-up approach, we conducted in-sample and out-sample tests. In the in-sample tests, we used different time ranges as training data, such as 2000-2010, 2001-2011, 2002-2012, and so on up to 2009-2019. We tested the model's performance by using the the remaining year data as test data. Figure 22 shows the weights for each asset class and signal during these tests. Over time, all trend signal values show a decreasing trend. For instance, Trend4 decreased from 0.52 in 2000-2010 to 0.41 in 2009-2019. This indicates that the strength of market trend signals is gradually weakening over time. Trend32 and Trend64 have zero values across all periods, suggesting that these signals did not contribute significantly during these test periods. Carry signals generally show an increasing trend, particularly Carry20 and Carry60, whose values gradually increase over different periods. For example, Carry20 increased from 0.07 in 2000-2010 to 0.17 in 2009-2019, while Carry60 increased from 0.24 in 2000-2010 to 0.42 in 2009-2019. However, Carry120 shows a decreasing trend, from 0.08 in 2000-2010 to nearly zero in 2009-2019. This indicates that market returns over longer holding periods are weakening. Agricultural and metals show variable trends, with some periods showing higher weights, indicating significance in those periods. Bonds, currencies, and equities have generally consistent weights, indicating their steady contribution over time. Energies show a noticeable weight in the earlier periods but decrease in significance in later periods.

Time Period	Trend2	Trend4	Trend8	Trend16	Trend32	Trend64	Carry5	Carry20	Carry60	Carry120	Agricultural Products	Bonds	Currencies	Energies	Equities	Metals
2000-2010	0	0.5221	0	0.0046	0.081	0	0	0.0673	0.2416	0.0833	0.0483	0.2355	0.4504	0.0851	0.1254	0.0553
2001-2011	0	0.6013	0	0.0172	0.0244	0	0	0.0567	0.2813	0.0191	0.0618	0.2301	0.4525	0.0745	0.1222	0.0588
2002-2012	0	0.5491	0	0.0093	0.0027	0	0	0.0625	0.3496	0.0267	0.0599	0.2052	0.4423	0.0749	0.1343	0.0834
2003-2013	0	0.5116	0	0.0131	0.0003	0	0	0.0683	0.3822	0.0246	0.0605	0.2072	0.4097	0.0703	0.1603	0.0919
2004-2014	0	0.4811	0	0.0174	0	0	0	0.0951	0.3978	0.0085	0.0604	0.2134	0.377	0.0665	0.1776	0.1052
2005-2015	0	0.4542	0	0.013	0	0	0	0.1039	0.4213	0.0076	0.0689	0.2264	0.3672	0.0596	0.1767	0.1011
2006-2016	0	0.4303	0	0.0031	0	0	0	0.1278	0.4322	0.0065	0.0798	0.223	0.3613	0.0561	0.1795	0.1003
2007-2017	0	0.4166	0	0.0002	0	0	0	0.1526	0.4274	0.0033	0.0726	0.2177	0.3685	0.0692	0.1845	0.0875
2008-2018	0	0.4172	0	0	0	0	0	0.1805	0.4015	0.0008	0.0799	0.2172	0.3586	0.0751	0.186	0.0833
2009-2019	0	0.413	0	0	0	0	0	0.1693	0.4177	0	0.0891	0.2114	0.319	0.0707	0.1953	0.1145

Figure 22: Cumulative returns of SG CTA and the bottom-up method

Figure 23,24, and 25 display the cumulative returns using the training data from 2000-2010, 2004-2014, and 2009-2019. Using earlier training data (2000-2010 and 2004-2014), the model demonstrates higher stability, with predicted results closely matching actual outcomes, showcasing the model's robustness. The model better learns the long-term trends of the market, thereby exhibiting strong predictive capabilities during the testing phase. However, when using more recent training data (2009-2019), the model's stability significantly decreases, with a larger discrepancy between predicted and actual results, indicating the model's inadequate adaptability to recent market changes.

For the out-of-sample tests, we evaluated how well the bottom-up method replicated other CTA funds using return data from Capital [2024a], Beacon [2024], Fund [2024], AlphaSimplex [2024], Capital [2024b], Fidelity [2024], and PIMCO [2024].

Figure 26 shows the weights for different asset classes and signals when replicating these

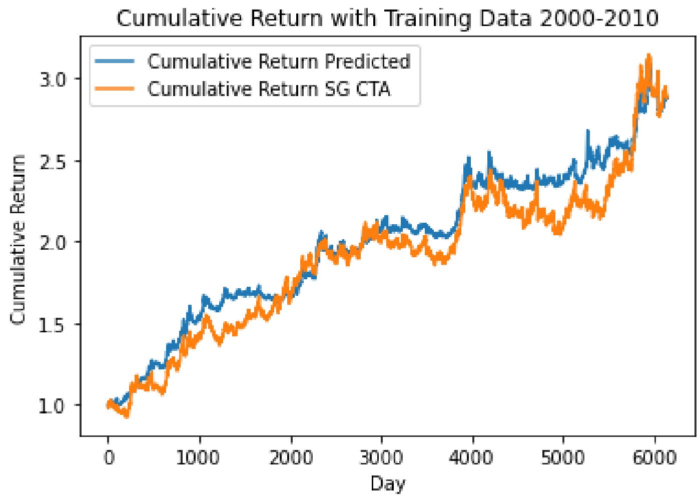

Figure 23: Cumulative returns of SG CTA and the bottom-up method

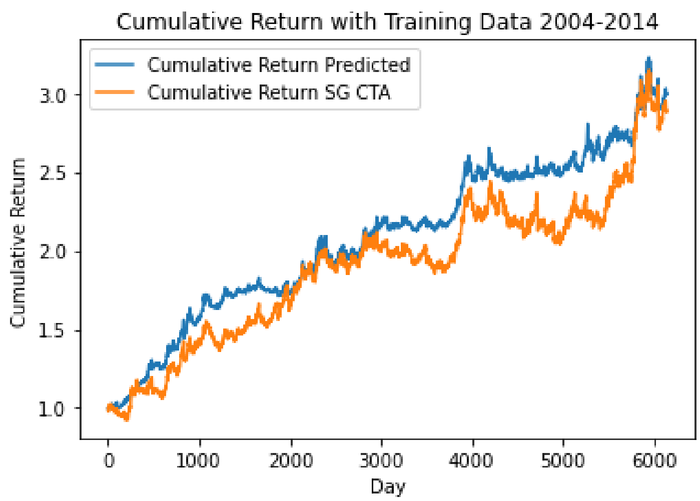

Figure 24: Cumulative returns of SG CTA and the bottom-up method

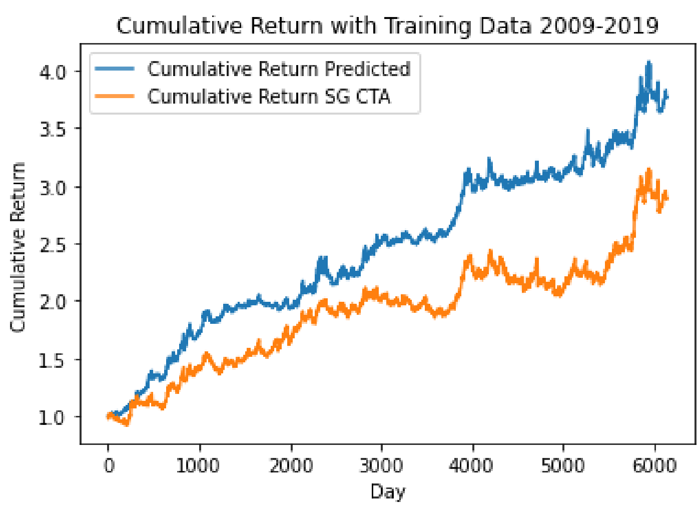

Figure 25: Cumulative returns of SG CTA and the bottom-up method

	Trend2	Trend4	Trend8	Trend16	Trend32	Trend64	Carry5	Carry20	Carry60	Carry120	Agricultural Products	Bonds	Currencies	Energies	Equities	Metals
ABYIX	0.0255	0.2148	0.0018	0.0113	0.2826	0.0018	0.0629	0.0271	0.3283	0.0582	0.1333	0.1819	0.3012	0.0966	0.2006	0.0866
AHLYX	0.0602	0.1403	0.0408	0.0288	0.3302	0.0767	0.0232	0.0744	0.2324	0.0006	0.0761	0.1569	0.3063	0.1081	0.2253	0.1267
AQMIX	0.0311	0.1703	0.0403	0.009	0.3242	0.0079	0.0374	0.0398	0.1989	0.0129	0.1982	0.2082	0.2746	0.116	0.1637	0.1082
ASFYX	0.0066	0.1578	0.0008	0.0148	0.2963	0.048	0.3021	0.048	0.4188	0.0003	0.2507	0.0501	0.2494	0.0828	0.2289	0.0723
EQCHX	0.0161	0.0066	0.0066	0.0271	0.2445	0.0245	0.0025	0.1167	0.0561	0.1115	0.2112	0.1115	0.2252	0.0557	0.2561	0.115
MFTFX	0.0033	0.0602	0.0001	0.0026	0.3126	0.002	0.6094	0.0241	0.1242	0.2242	0.2222	0.0113	0.1297	0.1383	0.2155	0.0523
PQTIX	0.0431	0.3703	0.0015	0.005	0.2893	0.0029	0.0936	0.0923	0.1263	0.2229	0.3227	0.2971	0.3557	0.097	0.0778	0.0443

Figure 26: Cumulative returns of SG CTA and the bottom-up method

CTA funds using the bottom-up method. Analysis reveals that Trend4, Carry60, Trend32, Carry120, and Carry20 have higher weights, indicating these signals are more significant in capturing market changes. Specifically, PQTIX and ABYIX have higher weights in short-term trend signals (Trend4), while AQMIX and AHLYX have higher weights in medium-term trend signals (Trend32). Conversely, MFTFX and EQCHX have lower weights in short-term trend signals (Trend2), and ABYIX and AHLYX also have lower weights in long-term trend signals (Trend64). In terms of carry signals, EQCHX has the highest weights in medium-term carry signals (Carry60), while PQTIX and AHLYX have higher weights in short-term carry signals (Carry20). However, EQCHX has lower weights in long-term carry signals (Carry120), and AHLYX and MFTFX also have lower weights in short-term carry signals (Carry5). For asset class weights, MFTFX has higher weights in the agriculture asset class, PQTIX and AHLYX have higher weights in the currencies asset class, and EQCHX and MFTFX have higher weights in the equities asset class. Conversely, PQTIX and MFTFX have lower weights in the metals asset class, and MFTFX and EQCHX have lower weights in the energies asset class. These results indicate that different CTAs have varying weights in different signals and asset classes, validating the stability of the bottom-up method under out-of-sample conditions.

Figures 27 shows the results of using the bottom-up method to predict the cumulative re-

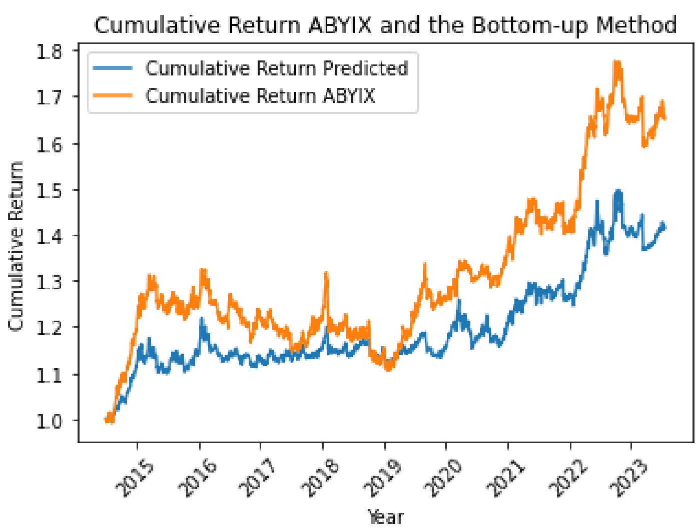

Figure 27: Cumulative returns of ABYIX and the bottom-up method

turns of ABYIX. The actual returns and predicted returns had a generally consistent trend from 2015 to early 2017, with some fluctuations but relatively small differences between them. However, from mid-2017 to mid-2019, the gap between the predicted returns and the actual returns significantly increased, indicating that the model's predictive performance was unsatisfying during this period. While the bottom-up method was able to track the actual ABYIX returns well in many time periods, its predictive performance notably declined during periods of significant market changes.

Figure 28 shows the cumulative return predictions for AQMIX using the bottom-up method.

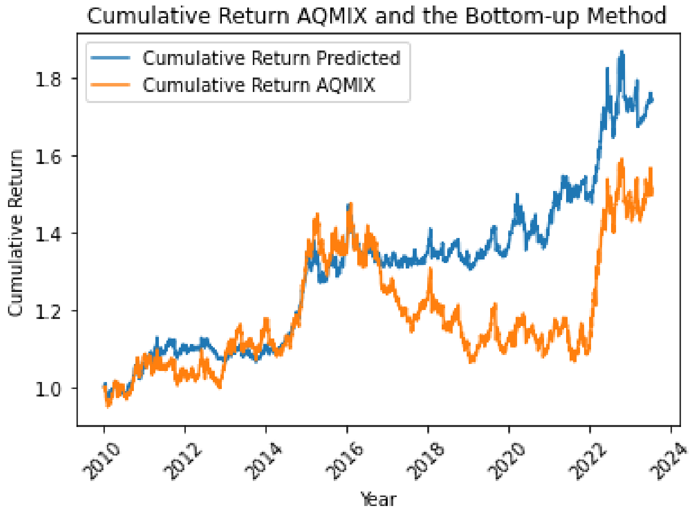

Figure 28: Cumulative returns of AQMIX and the bottom-up method

From 2010 to 2015, the model performs well, with actual and predicted returns aligning closely. However, in 2016, the model's accuracy declines significantly, showing large differences. Later, the model's performance improves again.

Figure 29 presents the cumulative return predictions for ASFYX using the bottom-up

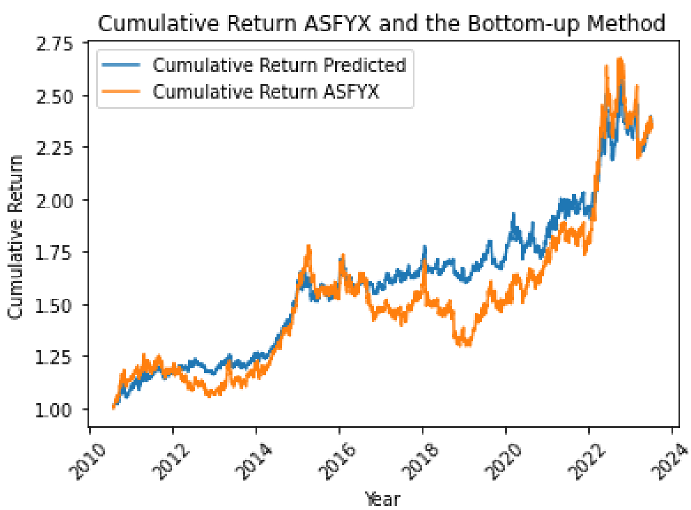

Figure 29: Cumulative returns of ASFYX and the bottom-up method

method. Between 2010 and 2016, both actual and predicted returns exhibit a generally consistent trend, marked by minor fluctuations and relatively small discrepancies. Post-2016, while the gap between the predicted and actual returns widened, the model still effectively captured market trends in most periods. Notably, from 2019 to early 2020, the predicted and actual returns converged once more, highlighting the model's short-term predictive strength. This suggests that the bottom-up method maintains a certain degree of stability in identifying basic market trends and is proficient in tracking the actual returns of ASFYX in most cases.

30 shows the cumulative return predictions for MFTFX using the bottom-up method. From 2015 to 2021, the actual returns and predicted returns align well, indicating that the model was effective in capturing market trends during this period. However, outside this period, the

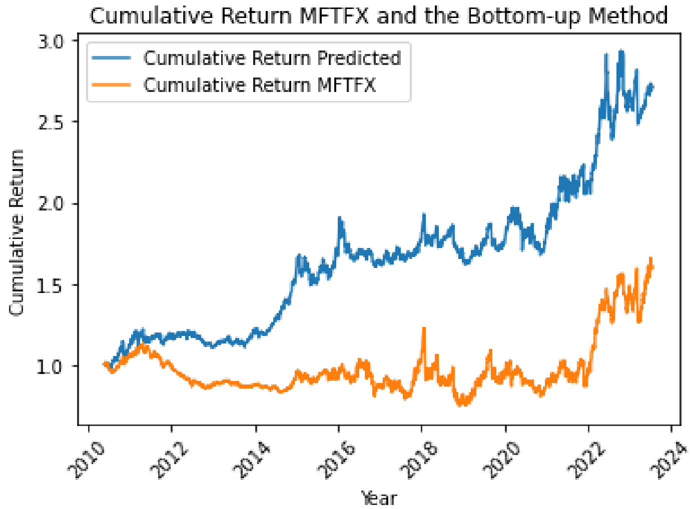

Figure 30: Cumulative returns of MFTFX and the bottom-up method

predicted returns diverged from the actual returns.

From 2019 onwards, the gap between actual and predicted returns significantly widened. The actual returns showed considerable volatility around 2020, which the models failed to predict accurately. Although actual returns recovered and grew after 2022, the predicted returns for all window sizes continued to decline, indicating persistent poor prediction performance. This demonstrates that the top-down method struggled to adapt to market changes and high volatility, regardless of the window size.

Figure 31 shows the cumulative return predictions for AHLYX using the bottom-up method.

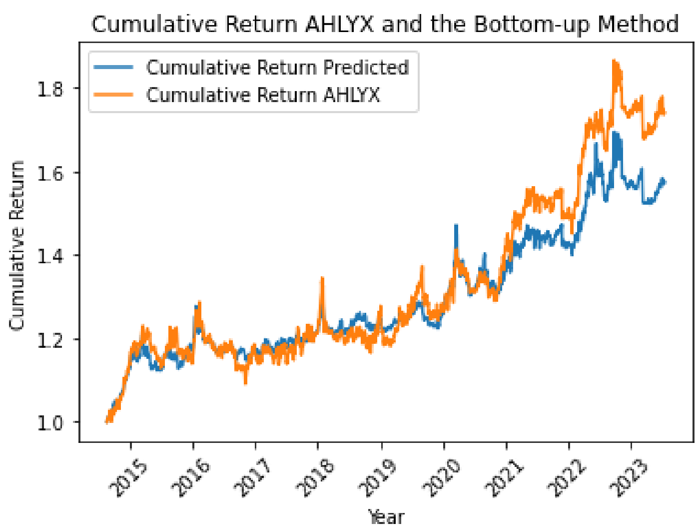

Figure 31: Cumulative returns of AHLYX and the bottom-up method

Except for 2021, the actual returns and predicted returns are generally consistent throughout most periods. Overall, the bottom-up method effectively tracks the market trends of AHLYX and provides accurate predictions in most cases.

Figure 32 shows the cumulative return predictions for EQCHX using the bottom-up method. Before 2015, the model's prediction accuracy was poor, with a large gap between actual and

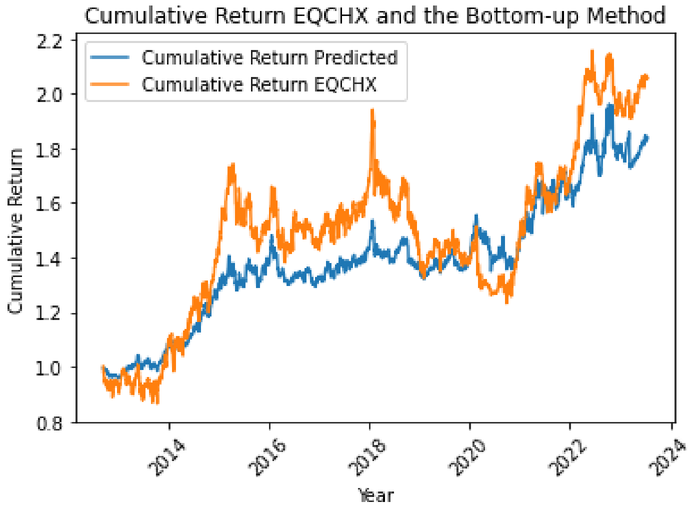

Figure 32: Cumulative returns of EQCHX and the bottom-up method

predicted returns. From 2015 to 2019, the actual and predicted returns were generally consistent, with small differences, demonstrating the model's effectiveness during this period.

From 2019 to 2022, the prediction accuracy declined again, with actual returns significantly higher than predicted returns. The model failed to capture the significant growth from 2020 to 2022 accurately. However, after 2022, the gap between actual and predicted returns narrowed, and the model generally tracked the actual returns well. Despite some prediction errors, the bottom-up method was overall effective in capturing the market trends of EQCHX.

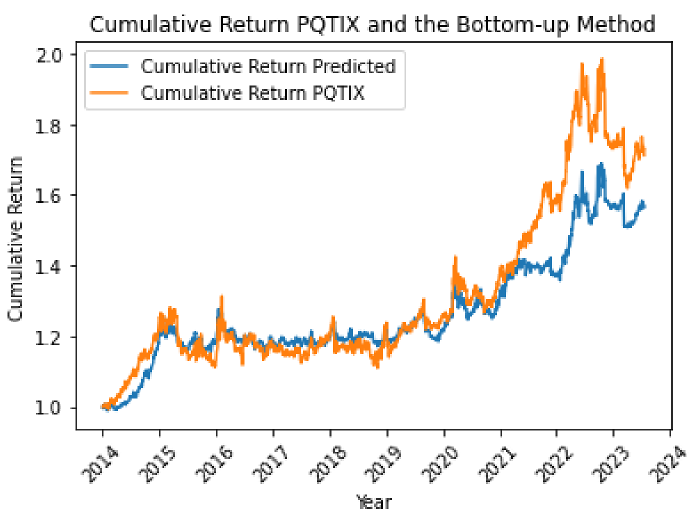

Figure 33: Cumulative returns of PQTIX and the bottom-up method

Figure 33 shows the cumulative return predictions for PQTIX using the bottom-up method. Apart from 2021, the actual returns and predicted returns are largely consistent throughout most periods. From 2014 to 2020, the model effectively captured market trends, with the actual and predicted returns displaying similar patterns and minimal differences, demonstrating the model's accuracy.

5.5 Comparison and analysis

The bottom-up method outperforms the top-down method in several aspects. Specifically, the top-down method's R^2 value is generally lower. For instance, when predicting using SG CTA data, the highest R^2 value achieved by the model is only 0.42, indicating that only 42% of the data can be explained by the model. Additionally, the top-down method's Sharpe ratio is low; in the same SG CTA data prediction, the best Sharpe ratio achieved is just 0.48, demonstrating poor risk-adjusted returns. In terms of robustness of the top-down method applied to other CTAs, though the method demonstrates some capability in capturing basic market trends, but it shows insufficient robustness in dealing with sudden market changes and disruptions. For example, in early 2016, there were significant discrepancies between the predicted and actual returns for ASFYX, EQCHX, AQMIX, AHLYX, and PQTIX.

In contrast, the bottom-up method excels in many areas. When predicting using SG CTA data, the bottom-up method achieves an R^2 value of 0.59, indicating that the model can explain 59% of the actual returns. The Sharpe ratio is also higher, reaching 0.80, indicating better risk-adjusted returns. Regarding robustness, the bottom-up method shows high stability in different time periods during in-sample tests. Although its predictions using later years as the training set for SG CTA are not as good as those using earlier years, it generally captures market trends accurately. In out-of-sample tests, the bottom-up method demonstrates higher robustness in adapting to market changes than the top-down method. Overall, in terms of accuracy, profitability, and robustness, the bottom-up method outperforms the top-down method.

6 Limitations and Future work

Despite the satisfactory results achieved using the top-down and bottom-up approaches, there are areas for discussion and potential enhancement. This section has a discussion of the limitations of this work and directions of future work.

6.1 Limitations

6.1.1 Trading costs and buffer

In the present work, for the sake of facilitating a more straightforward replication of CTA strategies, trading costs were not incorporated. However, in real-world trading scenarios, these costs play a pivotal role in shaping strategy decisions. A strategy that leads to frequent trading, despite being profitable in theory, might result in diminished actual returns due to high trading costs. This dynamic is a critical aspect that the current model does not capture.

The influence of trading costs cannot be understated. In many cases, they can significantly alter the profitability of a strategy. For instance, a strategy with high transaction frequency might appear profitable on paper, but when trading costs are factored in, the net returns could be substantially lower. This reality presents a major limitation in the current research, as it overlooks a crucial component that affects the practical viability of CTA strategies.

CTAs typically operate in highly liquid markets and trade at a slower pace, making the omission of trading costs more justifiable. For instance, strategies like Trend32, Trend64, Carry60, and Carry120 generally incur minimal trading costs in the most liquid markets. However, for quicker strategies (especially Trend2 and Trend4) and less liquid markets, incorporating trading costs is crucial for achieving realistic replication.

To address this limitation, future models can introduce the concept of buffering. Buffering entails not initiating a trade unless the forecasted change exceeds a certain threshold. A better analysis would also simply estimate trading costs and include them in the replications. Buffering operates on the principle that minor fluctuations in forecasts should not trigger immediate trading responses. By setting a predefined threshold for forecast changes, trades are only executed when the forecasted change is substantial enough to justify the transaction, thereby potentially increasing the net profitability of a strategy after accounting for trading costs.

6.1.2 Scope of the bottom-up method

The study predominantly focuses on trend following and carry strategies, which, while shows good results, do not encompass the full range of strategies employed by CTAs. To improve the results, we can expand into diverse strategies by incorporating a wider array of strategies, such as global macro strategies, the impact of regulatory changes, and sector-specific trades. It will provide a more comprehensive understanding of CTA activities.

6.1.3 Quantification of Uncertainty in Strategy Reconstructions

The absence of a quantification mechanism for the uncertainty in strategy reconstructions is a notable limitation. Incorporating uncertainty quantification is essential for comprehensive risk management and setting realistic expectations regarding model reliability. Future models could adopt statistical measures like confidence intervals or Bayesian methods to assess and communicate the inherent uncertainty in the reconstructions.

6.2 Future work

6.2.1 Extreme values in the bottom-up method

The results from the bottom-up replication were very extreme, with Trend4 and Carry60 having a disproportionately large impact, while Trend2, Trend8, Trend64, and Carry5 had almost no impact. Future work should thoroughly investigate the causes of these extreme results. Ensuring that the input data is accurate, complete, and appropriately preprocessed is crucial, as any anomalies or inconsistencies in the historical price data could significantly impact the outcomes. Reviewing and potentially adjusting the parameters used in the trend-following and carry strategies is another important step. Sensitivity analysis could help understand how different parameter values affect the results.

6.2.2 Differences between the results of the top-down method and DBMF

The replications in this study did not perform as well as DBMF claim to have achieved with their replication strategies. It would be valuable for future research to investigate the reasons behind this discrepancy. Analyzing the detailed methodologies and statistical models used by DBMF could provide insights into potential improvements for the current replication methods.

6.2.3 Combination of the bottom-up and top-down methods

The current research has individually explored the bottom-up and top-down methods for replicating CTA strategies, each yielding promising results. The intriguing question arises: can combining these two methods enhance the replication results? A hybrid approach that integrates the granular detail of the bottom-up method with a broader perspective of the top-down approach could potentially offer a more comprehensive and accurate replication of CTA strategies.

6.2.4 Investigation of Long-Only Biases

Some CTAs might be incorporating long-only biases in their strategies, subtly blending them with traditional trend and carry signals. Investigating this hypothesis is important to ascertain if such biases exist and how they influence overall strategy performance and risk profiles. Uncovering these biases would reveal new facets in CTA strategies, offering insights into hidden risk exposures and portfolio management tactics.

References

AlphaSimplex. **ASFYX**. https://www.alphasimplex.com/products/managed-futures-strategy, 2024. Accessed on January 9th, 2024.

Noël Amenc, Lionel Martellini, Jean-Christophe Meyfredi, and Volker Ziemann. Passive hedge fund replication–beyond the linear case. *European Financial Management*, 16(2):191–210, 2010.

Clifford S Asness, Tobias J Moskowitz, and Lasse Heje Pedersen. Value and momentum everywhere. *The journal of finance*, 68(3):929–985, 2013.

American Beacon. **AHLYX**. https://www.americanbeaconfunds.com/mutual_funds/AHL.aspx, 2024. Accessed on January 9th, 2024.

Geetesh Bhardwaj, Gary B. Gorton, and K. Geert Rouwenhorst. Fooling Some of the People All of the Time: The Inefficient Performance and Persistence of Commodity Trading Advisors. *The Review of Financial Studies*, 27(11):3099–3132, 07 2014. ISSN 0893-9454. doi: 10.1093/rfs/hhu040. URL https://doi.org/10.1093/rfs/hhu040.

Abbey Capital. **ABYIX**. https://www.abbeycapital.com/abbey-capital-futures-strategy-fund/, 2024a. Accessed on January 9th, 2024.

Chesapeake Capital. **EQCHX**. https://www.axsinvestments.com/axs-chesapeake-strategy-fund/, 2024b. Accessed on January 9th, 2024.

Robert Carver. *Advanced Futures Trading Strategies: 30 fully tested strategies for multiple trading styles and time frames*. Harriman House Limited, 2023a.

Rovert Carver. **Adjusted future price**. https://github.com/robcarver17/pysystemtrade/tree/master/data/futures/adjusted_prices_csv, 2023b. Accessed on July 12, 2023.

DBMF. **iMGP DBi managed futures strategy ETF**. https://imgpfunds.com/im-dbi-managed-futures-strategy-etf/, 2013. Accessed on July 11, 2023.

Fidelity. **MFTFX**. https://fundresearch.fidelity.com/mutual-funds/ratings/042765875, 2024. Accessed on January 9th, 2024.

AQR Managed Futures Strategy Fund. **AQMIX**. https://funds.aqr.com/funds/alternatives/aqr-managed-futures-strategy-fund/aqmix, 2024. Accessed on January 9th, 2024.

William Fung and David A Hsieh. The information content of performance track records: investment style and survivorship bias in the historical returns of commodity trading advisors. *Journal of Portfolio Management*, 24(1):30–41, 1997.

Greg N Gregoriou and Yao Chen. Evaluation of commodity trading advisors using fixed and variable and benchmark models. *Annals of Operations Research*, 145:183–200, 2006.

The Guardian. Top 10 hedge funds made £1.5bn profit from ukraine war food price spike. https://www.theguardian.com/world/2023/apr/14/hedge-funds-profit-ukraine-war-food-price-surge, 2023. Accessed on July 11, 2023.

Jasmina Hasanhodzic and Andrew W Lo. Can hedge-fund returns be replicated?: The linear case. *The Linear Case (August 16, 2006)*, 2006.

Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91, 1993.

Narasimhan Jegadeesh and Sheridan Titman. Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of finance*, 56(2):699–720, 2001.

Narasimhan Jegadeesh and Sheridan Titman. Cross-sectional and time-series determinants of momentum returns. *The Review of Financial Studies*, 15(1):143–157, 2002.

Ralph SJ Koijen, Tobias J Moskowitz, Lasse Heje Pedersen, and Evert B Vrugt. Carry. *Journal of Financial Economics*, 127(2):197–225, 2018.

Tobias J Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. Time series momentum. *Journal of financial economics*, 104(2):228–250, 2012.

PIMCO. **PQTIX**. https://www.pimco.com/en-us/investments/mutual-funds/trends-managed-futures-strategy-fund/inst, 2024. Accessed on January 9th, 2024.

Barclays Bank PLC. **SG CTA Index**. https://portal.barclayhedge.com/cgi-bin/indices/displayHfIndex.cgi?indexCat=SG-Prime-Services-Indices&indexName=SG-CTA-Index, 2023. Accessed on July 11, 2023.

Akihiko Takahashi, Kyo Yamamoto, et al. Hedge fund replication. *The Recent Trend of Hedge Fund Strategies*, pages 57–95, 2008.