Probabilistic Portfolio Strategy Using Hidden Markov Models

A Walk-Forward Strategy for
Risk-Aware Allocation Across SPY and GLD

Author: *Christian Tocco*

Project Timeline: *February 2025 - July 2025*

GitHub Repository: https://github.com/christian-tocco/HMM-portfolio-strategy.git

Table of Contents:

1. Executive Summary

This report explores the development and evaluation of a systematic trading strategy based on **Hidden Markov Models** (HMMs) applied to financial time series data.

The objective is to construct a regime-aware asset allocation model capable of dynamically adjusting portfolio weights in response to changing market conditions.
The strategy allocates capital weekly between U.S. equities (***SPY***) and gold (***GLD***), using probabilistic state inference and volatility-based indicators to identify risk regimes.
The model employs *Gaussian Mixture Hidden Markov Models* to infer latent states from observed returns and volatility, and applies walk-forward optimization to preserve out-of-sample validity and avoid lookahead bias.

Two strategy variants are developed:

- An optimized version focusing on *return maximization* through probabilistic risk allocation;
- A conservative version that integrates risk management filters based on recent volatility and drawdown behavior.

The strategies are benchmarked against passive alternatives, including the S&P 500 Total Return Index, GLD, and a 50/50 static allocation, to assess both return and risk-adjusted performance over a **20**-year backtest horizon.

By combining statistical modeling with real-world trading logic, I aimed to explore how data-driven decision-making can improve long-term portfolio outcomes and help bridge the gap between theoretical models and practical asset management.

2. Theoretical Background

In financial markets, asset returns are often influenced by underlying regimes such as bull and bear markets, high or low volatility periods, or monetary policy cycles.
Traditional models, however, generally assume constant statistical properties over time, an assumption which often fails in practice.
This motivates the use of *regime-switching models*, where asset behavior is modeled as a function of latent (unobservable) states.

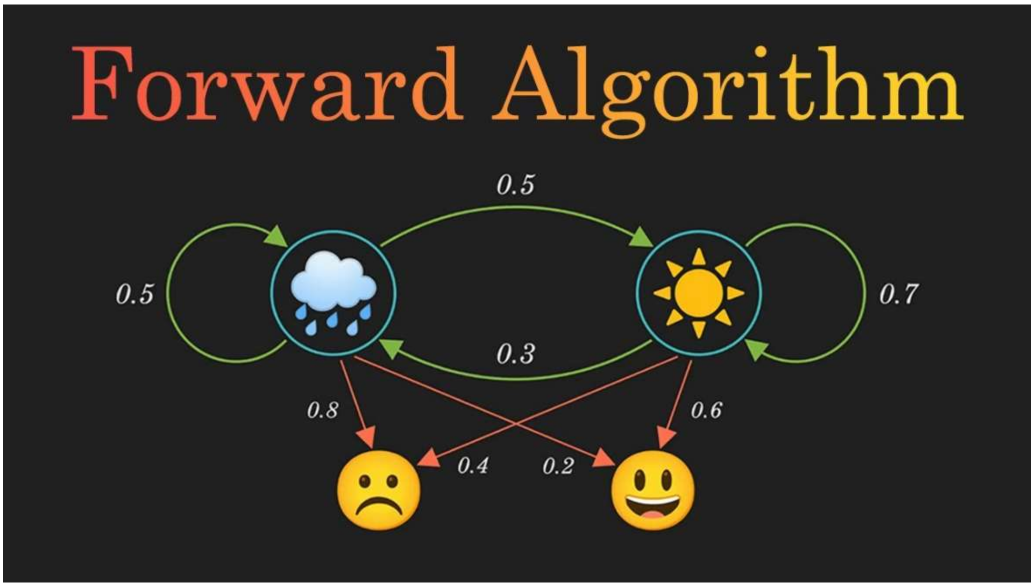
Hidden Markov Models provide a statistical framework to model such behavior.
An HMM assumes that:

1. The market evolves through a sequence of hidden states (ex, **risk-on, risk-off**).
2. Each hidden state generates observable variables (ex, **returns**, **volatility**) according to a specific probability distribution.
3. State transitions follow a Markov process, meaning the next state depends only on the current state.

In a **Gaussian Mixture HMM**, each hidden state emits observations drawn from a mixture of Gaussian distributions, allowing for richer modeling of asset return distributions and regime complexity.
This makes GMMHMM particularly well-suited for financial data, which often exhibits ***non-normality, volatility clustering,*** and ***heavy tails***.

By applying HMMs to time series of asset returns and derived indicators (ex, *volatility*), it is possible to infer the probability of being in each market regime, and use this information to adjust portfolio allocations dynamically.

This approach contrasts with traditional models like CAPM or static risk parity, which assume **fixed relationships** over time. HMM-based strategies instead adapt continuously, making them appealing for tactical allocation, risk mitigation, and alpha generation.



3. Data and Features

The strategy is built using publicly available financial market data covering the period from ***2005*** to ***2025***, with a *weekly frequency* (Friday close).
The two primary asset classes considered for allocation are:

- **SPY** - the S&P 500 ETF, used as a proxy for U.S. equity exposure;
- **GLD** - the Gold Shares ETF, representing exposure to gold as a *defensive asset*.

In addition to price-based returns, the analysis also incorporates **SPY Total Return** data, which includes dividend reinvestment and serves as a more accurate benchmark for long-term investment performance.

3.1 Return Calculations

All returns are computed using logarithmic returns, a standard approach due to their time-additive properties and symmetry:

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

Returns are resampled to weekly frequency using the last available closing price of each week.

3.2 Volatility Feature

A key component of the model input is a *rolling volatility estimate*, calculated as the 4-week standard deviation of SPY log returns.
This feature serves as a proxy for recent market risk and is used as an input to the Hidden Markov Model to assist with regime classification.

$$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n-1}(r_{t-i} - \overline{r})^2}$$

where $n = 4$ weeks and $\overline{r}$ is the mean return over the window.

3.3 Combined Dataset

The final modeling dataset includes the following columns:

- SPY_Return: Weekly log return of SPY
- GLD_Return: Weekly log return of GLD
- SPY_TR_Return: Weekly log return of SPY Total Return index
- Volatility: 4-week rolling standard deviation of SPY returns

These features are aligned and merged using inner joins on the weekly timestamps, ensuring consistency across all asset return series.

3.4 Sample Data Snapshot

The table below illustrates a sample of the weekly data used for model training and strategy execution:

Date	SPY_Return	GLD_Return	SPY_TR_Return	Volatility
2007-01-05	0.0052	-0.0024	0.0054	0.0061
2007-01-12	-0.0011	0.0046	-0.0009	0.0058
2007-01-19	0.0033	-0.0018	0.0036	0.0054
2007-01-26	0.0020	0.0065	0.0022	0.0047
2007-02-02	-0.0007	-0.0031	-0.0004	0.0045

Note: Returns are log-transformed and resampled weekly (Friday close).

3.5 Data Sources

- *SPY* and *GLD* historical price data were downloaded directly from **Interactive Brokers** through the Trader Workstation API.
- *SPY Total Return* data was obtained from **Yahoo Finance** via the yfinance Python package, using the ticker ^SP500TR.

All data were pre-processed in Python using the *Pandas* and *NumPy* libraries, and aligned to a weekly frequency using the *'W-FRI'* resampling convention to reflect realistic execution timing.

4. Methodology

The core of the strategy lies in the application of a **Hidden Markov Model** to detect latent market regimes and allocate capital probabilistically between SPY and GLD. The model is retrained weekly using a walk-forward optimization framework, ensuring that each allocation decision is made based solely on information available at the time.

4.1 Model Framework

The strategy is based on the following steps, repeated each week in a walk-forward fashion:

1. *Data Preparation*
   A rolling window of **104** weeks (*2 years*) of past data is extracted, including:
   - Weekly log returns of SPY (SPY_Return).
   - 4-week rolling volatility of SPY (Volatility).
2. *Feature Scaling*
   Features are standardized using a *StandardScaler* to normalize the input and improve model convergence.
3. *Model Training*
   A **Gaussian Mixture Hidden Markov Model** is trained on the scaled data.

The model includes:

- n_components = 2 hidden states
- n_mix = 3 Gaussian mixtures per state
- Covariance type: tied
- Maximum iterations: 100

4. *State Inference and Transition Prediction*

   The model produces:

   - Posterior probabilities for each state at time t.
   - Forecast of next-period state probabilities $\pi_{t+1} = \pi_t \cdot A$, where A is the transition matrix.

5. *Risk-Based Allocation*
   Each hidden state is evaluated based on its historical **Sharpe Ratio**.
   The state with the lowest Sharpe is considered "high risk." The portfolio weight for GLD is then set equal to the probability of being in that high-risk state at $t+1$.

$$\omega_{GLD} = \pi_{t+1}^{(high\,risk\,state)} \quad , \quad \omega_{SPY} = 1 - \omega_{GLD}$$

6. *Fallback Rule*
   If model training fails (ex, convergence error), the allocation defaults to **50/50**.

4.2 Strategy Variants

Two versions of the strategy were implemented:

- **Optimized Version:**
  A pure HMM-based probabilistic allocation using the procedure above, without any manual intervention or constraint.
- **Conservative Version:**
  Builds upon the optimized strategy by introducing two layers of risk control filters:
  - *Soft filter*: If current volatility exceeds **2.5%**, SPY exposure is capped at 70%
  - *Hard filter*: If SPY experiences a drawdown greater than -15% over the past 12 weeks, the model allocates 100% to GLD

These filters aim to mitigate risk during periods of heightened uncertainty or sustained market stress, enhancing the strategy's robustness.

4.3 Method summary

The diagram summarizes the weekly walk-forward process implemented in the strategy.
After preparing and scaling the input data, a *GMMHMM* is trained to detect hidden regimes.
The model estimates transition probabilities and allocates capital based on the likelihood of entering a high-risk state.
If model convergence fails, a fallback allocation of *50/50* is applied.

5. Strategies Overview

Two strategy variants were developed from the core Hidden Markov Model framework, both operating under a weekly walk-forward regime. While they share the same probabilistic allocation logic and training methodology, they differ in terms of **risk management philosophy** and robustness under stress conditions.

5.1 Optimized Strategy (Pure HMM Allocation)

The optimized strategy focuses purely on **return maximization** by allocating capital based on the probabilistic likelihood of entering a high-risk regime.

Each week, the strategy:

- Trains a *GMMHMM* on the past 104 weeks of SPY return and volatility data
- Identifies the state with the lowest historical Sharpe Ratio (considered "**high-risk**")
- Allocates capital to GLD proportionally to the forecasted probability of being in that state
- Allocates the remaining capital to SPY.

This version does not apply any risk filters or override mechanisms.
Its strength lies in its flexibility and responsiveness to regime shifts, potentially capturing more upside during favorable market conditions.
However, it may be more exposed during extreme market stress (*more drawdown*).

5.2 Conservative Strategy (HMM + Risk Filters)

The conservative strategy builds on the same probabilistic core but incorporates additional layers of explicit **risk control**.

After computing the HMM-based allocation, two filters are applied sequentially:

1. **Volatility Filter** (Soft Constraint):
    - If the current 4-week rolling volatility exceeds 2.5%, SPY exposure is capped at 70%.
2. **Drawdown Filter** (Hard Constraint)**:**
    - If SPY has experienced a drawdown of more than -15% over the past 12 weeks, the model shifts fully to GLD (SPY weight = 0%).

These filters act as override conditions that limit equity exposure during heightened uncertainty or market corrections.
This makes the strategy more suitable for capital preservation and reduces tail-risk sensitivity.

5.3 Strategic Comparison

Aspect	Optimized Strategy	Conservative Strategy
Objective	*Return maximization*	*Drawdown control + stability*
Allocation logic	*Pure probabilistic*	*Probabilistic + override filters*
Risk filters	*None*	*Volatility + Drawdown*
Responsiveness	*High*	*Moderate*
Robustness in crisis	*Lower*	*Higher*
Implementation complexity	*Lower*	*Slightly higher*

6. Performance Evaluation

The two strategy variants were evaluated over a **20**-year backtest period (*2005 - 2025*) using a weekly walk-forward framework. All results are out-of-sample, generated without access to future data, and include realistic assumptions regarding data frequency and timing.

The performance of each strategy is benchmarked against:

- **SPY** - price returns (ex-dividends).
- **SPY TR** - total return including dividends.
- **GLD** - gold ETF returns.
- A static **50/50** portfolio **between SPY** and **GLD** (rebalanced weekly).

All strategies are assessed using the following metrics:

- *CAGR* (Compound Annual Growth Rate)
- *Max Drawdown* (worst peak-to-trough decline)
- *Average Drawdown* (average of all negative drawdown periods)
- *Sharpe Ratio* (risk-adjusted return using standard deviation)

6.1 Results Summary

Strategy	CAGR	Max Drawdown	Avg Drawdown	Sharpe Ratio
Optimized	*11.69%*	*-31.03%*	*-7.10%*	*0.7597*
Conservative	*10.68%*	*-22.88%*	*-7.65%*	*0.7424*
SPY	*6.51%*	*-59.66%*	*-14.78%*	*0.4313*
SPY TR	*8.66%*	*-58.16%*	*-12.87%*	*0.5416*
GLD	*7.39%*	*-48.05%*	*-22.87%*	*0.5093*
50/50 Portfolio	*8.57%*	*-29.73%*	*-5.19%*	*0.7022*

	CAGR	Max Drawdown	Average Drawdown	Sharpe Ratio
Strategy	0.1068	-0.2288	-0.0765	0.7424
SPY	0.0651	-0.5966	-0.1478	0.4313
SPY TR	0.0866	-0.5816	-0.1287	0.5416
GLD	0.0739	-0.4805	-0.2287	0.5093
50/50	0.0857	-0.2973	-0.0519	0.7022

Conservative Strategy Result computed

	CAGR	Max Drawdown	Average Drawdown	Sharpe Ratio
Strategy	0.1169	-0.3103	-0.0710	0.7597
SPY	0.0651	-0.5966	-0.1478	0.4313
SPY TR	0.0866	-0.5816	-0.1287	0.5416
GLD	0.0739	-0.4805	-0.2287	0.5093
50/50	0.0857	-0.2973	-0.0519	0.7022

Optimized Strategy Result computed

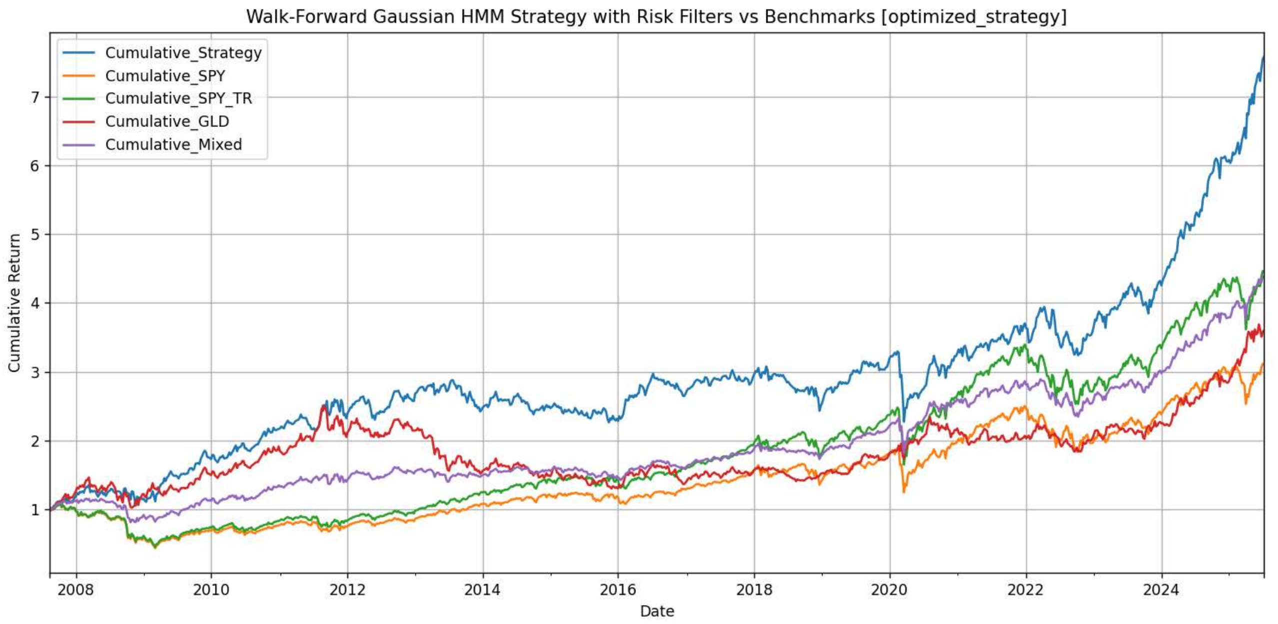
6.2 Key Observations

- Both strategy variants outperform **SPY** and **SPY TR** in terms of *CAGR* and *Sharpe Ratio*.
- The *Optimized Strategy* achieves the highest *CAGR* (**11.69%**), but with a higher drawdown.
- The *Conservative Strategy* delivers superior **drawdown control**, with Max Drawdown reduced to -22.88% compared to -59.66% for SPY.
- Risk-adjusted performance, as measured by the Sharpe Ratio, is significantly higher for both strategies than for any benchmark.

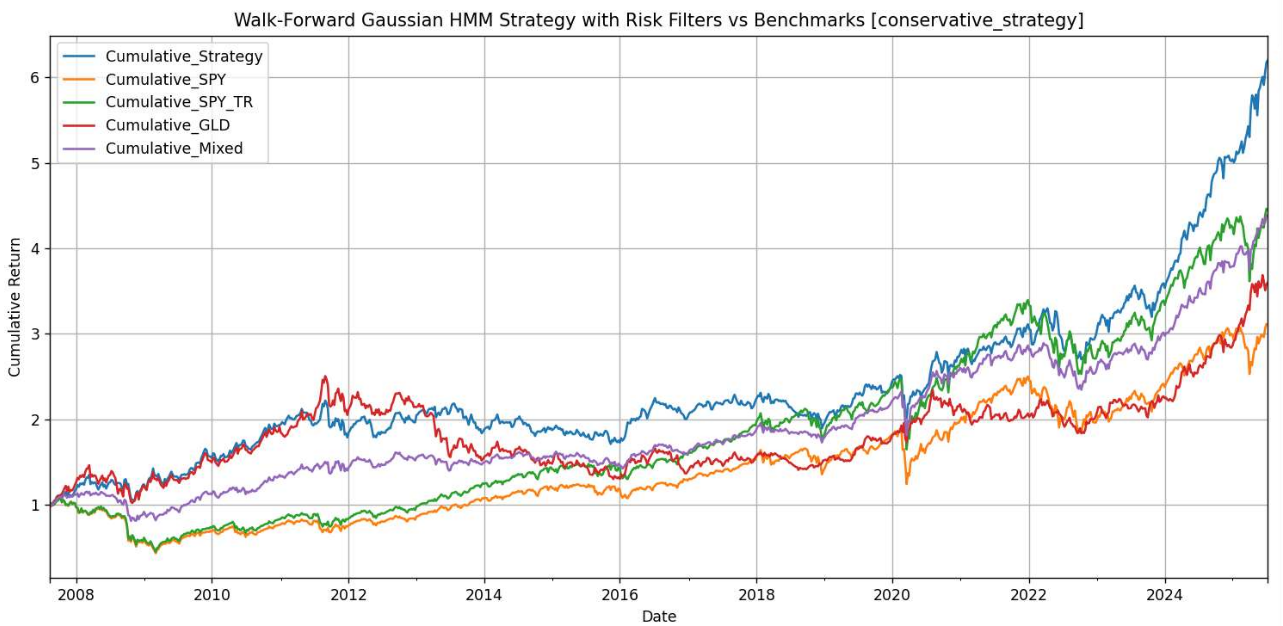
These results suggest that probabilistic regime-aware allocation, even in its simplest form, can lead to meaningful improvements in portfolio efficiency, particularly during market crises or periods of elevated volatility.

6.3 Visual Representation

The images below illustrate the cumulative performance of both strategies compared to the benchmarks over the full backtest horizon.



Cumulative returns of the optimized Hidden Markov Model strategy vs. benchmarks (2007 - 2025).



Cumulative returns of the conservative Hidden Markov Model strategy vs. benchmarks (2007 - 2025)

6.4 Interpretation and Risk of Overfitting

Both strategy variants exhibit a noticeable surge in performance starting from **late 2020 through 2025**, visibly outperforming the benchmarks during this period. While this may be a genuine consequence of structural changes in market behavior (ex, *elevated volatility*, *monetary tightening*, *macro uncertainty*), it also raises a key modeling concern:

- *How do we distinguish true model edge from overfitting to recent market dynamics?*

This question is central in quantitative strategy evaluation.
Several considerations help address this concern:

- The strategies were tested over a **20**-year horizon, which includes multiple market regimes: pre-crisis, GFC (2008), recovery, QE era, COVID crash, and post-2020 tightening.
- The walk-forward design with out-of-sample prediction helps mitigate lookahead bias.
- Performance during prior periods (**2008 - 2016**) was modest and consistent, showing no signs of curve-fitting or parameter instability.

Nevertheless, the post-2020 outperformance could be partially driven by:

- **GLD's defensive strength** during inflationary shocks.
- **Model sensitivity** to recent volatility patterns, which may not generalize indefinitely;
- A potential shift in the market structure that the model inadvertently exploited.

7. Benchmarking

Benchmarking provides essential context for evaluating the performance and risk-adjusted returns of the proposed Hidden Markov Model strategies.
In this section, we compare both the *optimized* and *conservative* variants against multiple baseline portfolios, including equity-only and blended approaches.

7.1 Selected Benchmarks

To ensure a fair and multi-dimensional evaluation, we selected the following benchmarks:

- **SPY:** Standard S&P 500 ETF, reflecting equity performance <u>without</u> dividends.
- **SPY Total Return:** Includes *dividends*, providing a more realistic long-term equity benchmark.
- **GLD:** Gold ETF, used as the primary *defensive asset* in our strategy.
- **50/50 Portfolio:** A static allocation with 50% SPY and 50% GLD, rebalanced weekly.

- Optimized Strategy:

Metric	Strategy	SPY	SPY TR	GLD	50/50
CAGR	*11.69%*	*6.51%*	*8.66%*	*7.39%*	*8.57%*
Max Drawdown	*-31.03%*	*-59.66%*	*-58.16%*	*-48.05%*	*-29.73%*
Sharpe Ratio	*0.76*	*0.43*	*0.54*	*0.51*	*0.70*

- Conservative Strategy:

Metric	Strategy	SPY	SPY TR	GLD	50/50
CAGR	*10.68%*	*6.51%*	*8.66%*	*7.39%*	*8.57%*
Max Drawdown	*-22.88%*	*-59.66%*	*-58.16%*	*-48.05%*	*-29.73%*
Sharpe Ratio	*0.74*	*0.43*	*0.54*	*0.51*	*0.70*

Insight: Both strategies significantly outperform *SPY* and *GLD* on a risk-adjusted basis. Even against the static 50/50 portfolio, the walk-forward HMM strategies show higher Sharpe Ratios and lower drawdowns, supporting their effectiveness in regime-switching environments.

8. Limitations and Robustness

While the presented strategies deliver compelling performance over a multi-decade horizon, it is critical to acknowledge their limitations and stress-test their robustness to ensure professional credibility and real-world applicability.

8.1 Data Limitations

- The strategy is based solely on weekly price returns of SPY and GLD. It does not incorporate *macro factors, options market data*, or *volume/liquidity signals* that could enrich the model's understanding of regime changes.

- SPY Total Return data was used only as a *performance benchmark*, not as an input feature, which may underrepresent dividend-adjusted market dynamics.

- Data is *clean and survivorship-free*, but still limited in scope to two assets, which can create concentration risks.

8.2 Model Assumptions

- The use of **Gaussian Mixture** Hidden Markov Models assumes stationarity of latent regimes, which may not hold over long periods, especially in structurally shifting markets.

- The number of **hidden states (**$n = 2$**)** and rolling window (**104** weeks) were selected via heuristic tuning. A more rigorous hyperparameter optimization could yield better generalization.

- In the optimized strategy, allocation weights are derived from **state-dependent Sharpe ratios**, which may amplify estimation noise in turbulent markets.

8.3 Implementation Risks

- The walk-forward setup approximates realistic rebalancing, but **transaction costs** are **not modeled**. Although annualized turnover can reach over 400% in certain periods and may reduce net returns, this effect is partially mitigated by the use of *highly liquid and low-cost ETFs* such as SPY and GLD, which are among the most cost-efficient instruments in their respective asset classes.

- The model occasionally fails to converge or assign meaningful state probabilities. These exceptions default to a *50/50 allocation*, which could dilute the model's signal strength.

8.4 Robustness Measures

Despite these limitations, several robustness features are integrated:

- *Walk-forward out-of-sample evaluation*, avoiding lookahead bias.
- **Soft** and **hard volatility/drawdown filters** in the conservative variant to protect capital during market stress.
- Strong outperformance during major market crises (ex, 2008, COVID-19) suggests regime detection is effective.
- Benchmark comparison with SPY, SPY TR, GLD, and a 50/50 mix ensures performance is not inflated by cherry-picking.

9. Conclusions and Future Work

This project demonstrates the feasibility and effectiveness of using *Hidden Markov Models* for regime-aware asset allocation. By leveraging probabilistic state inference and applying a walk-forward retraining structure, the proposed strategies achieved **outperformance** over traditional passive benchmarks such as *SPY* and *SPY Total Return*.

Two distinct implementations were explored:

- An *optimized version*, focused on return maximization through pure probabilistic allocation;
- A *conservative version*, enhanced with volatility and drawdown filters to reduce exposure during adverse conditions.

Both strategies delivered *higher Sharpe Ratios* and lower drawdowns compared to SPY, GLD, and a static 50/50 portfolio, while preserving consistency across various market regimes.

Moreover, the use of **transparent, low-cost ETFs**, combined with a disciplined walk-forward process, ensures practical implementability for real-world applications.

9.1 Future Work

Several directions may be explored to further refine and expand the framework:

- *Distributional Robustness*: Replace Gaussian mixtures with heavy-tailed distributions (ex, **Student-t**) to better model extreme events and fat-tailed returns.
- *Macro Regime Integration*: Incorporate macroeconomic indicators (ex, VIX, interest rates, yield curve slope) to inform or adjust regime probabilities.
- *Multi-Asset Extension*: Expand the strategy to include additional asset classes such as bonds or commodities, enabling more diversified allocation.
- *Ensemble Models*: Combine *multiple HMMs* or integrate them with machine learning classifiers (ex, Random Forests, LSTM) for more stable signal generation.
- *Dynamic Rebalancing Thresholds*: Optimize trading frequency and introduce adaptive thresholds to **reduce transaction costs**.
- *Live Deployment*: Transition the strategy to paper trading or simulated execution using APIs (ex, Interactive Brokers) for real-time validation.

These extensions would enhance both the robustness and adaptability of the framework, offering a more complete foundation for real-world applications in evolving market conditions.

All code, datasets, backtesting results and plotting scripts are available in the public GitHub repository linked at the beginning of this report.