

Regime Detection and Risk Allocation Using Hidden Markov Models

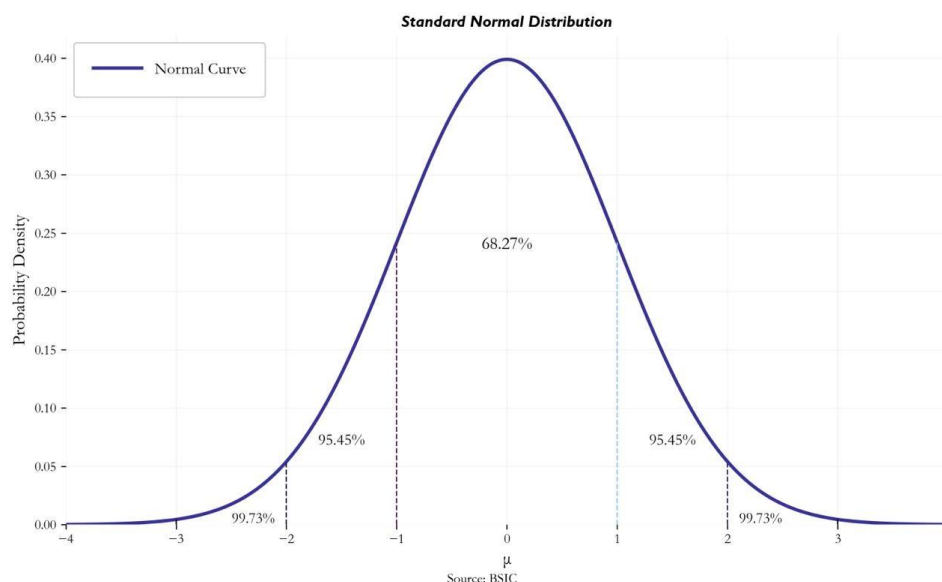
Introduction

Markets have long been known to exhibit persistent features over a period of time, through the years they have been described in several different ways: calm and turbulent, bullish and bearish, liquid and illiquid, due to macro conditions, geopolitical events, tech disruption, etc. These terms all refer to qualitative considerations on the current external environment. From a quantitative point of view this premise specifies that markets tend to follow certain statistical properties that persist over a period of days, months or even years, certainly contrasting the traditional model's views that returns are drawn from a predefined, stable probability distribution. For this reason practitioners increasingly have turned to Hidden Markov Models (HMMs).

In this article we start by describing the building blocks of HMMs: explaining the use of markov chains, defining latent processes underlying an HMM and connecting the theory to actual implementation to build trading strategies.

Markov Chains

In the “finance space”, returns are modeled as random variables drawn from a predefined probability distribution. The objective is not to debate the randomness itself, but to give a structure that governs the likelihood of different states of the world.



For example, if we use the normal distribution $\mathcal{X} \sim N(\mu, \sigma^2)$ to describe the different states of the world of NVIDIA's returns, we would notice that if this were the “true” distribution, every return above or below 4 standard deviations would take thousands of years to observe, while we know this is not the case. This model is not capturing changes in volatility, macro regimes, or other factors that influence the dynamics of the stock return, and by doing so we are underestimating risk in the tails and producing inaccurate estimates of the likelihood of various outcomes.

It is reasonable to think that at every point in time the probability distribution underlying the stock return is going to change; for example if there is high vol, negative returns are far more likely than positive return. Therefore we

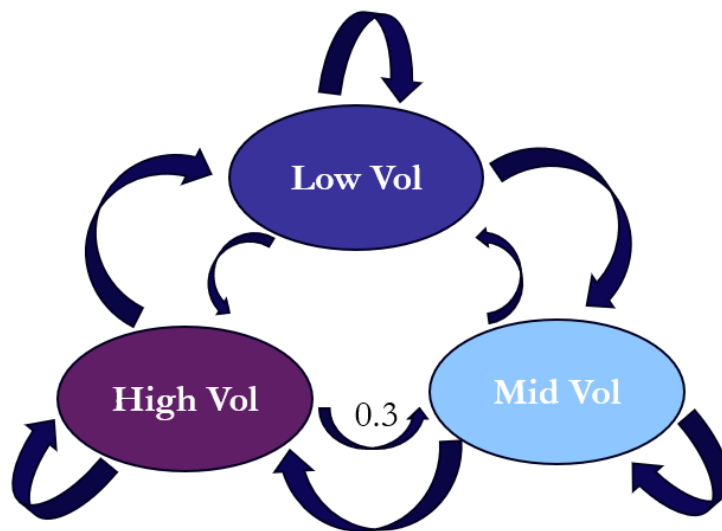
All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

say that the shape of our distribution is going to be governed by these “latent processes” (vol, trend, etc), and in order to model these evolving dynamics we rely on Markov Chains.

A Markov chain is a model that describes the probabilities of sequences of random variables, “states”, each of which can take on values from some set. A Markov chain makes a very strong assumption that if we want to predict the future in the sequence, all that matters is the current state.

Mathematically: $P(X_t, X_{t-1}, \dots, X_1) = P(X_{t+1} | X_t)$

Here is a diagram to better understand the idea of a Markov Chain...



Source: BSIC

The nodes labeled *High Vol*, *Low Vol*, and *Med Vol*, corresponds to the states $Q=\{q_1, q_2, q_3\}$ of the chain. The arrows, each labeled with the probability of moving from state i to state j are going to form the transition probability matrix $A=[a_{ij}]$. For example, the arrow from *High Vol* to *Med Vol* with 0.3 means $a_{13}=0.3$. In this way, every row of A represents one state and lists the probabilities of transitioning from that state to all others and the values in each row always sum to 1. $\sum_{j=1}^N a_{ij} = 1$.

In practice, the transition probabilities in the matrix A are estimated using Maximum Likelihood Estimation (MLE). The idea is that the model selects the set of parameters that maximize the likelihood of observing the given sequence of states and for a standard Markov Chain where states are observable, this simply means counting transitions: $p_{ij} = \frac{\text{number of transitions from state } i \text{ to state } j}{\text{total transitions from state } i}$

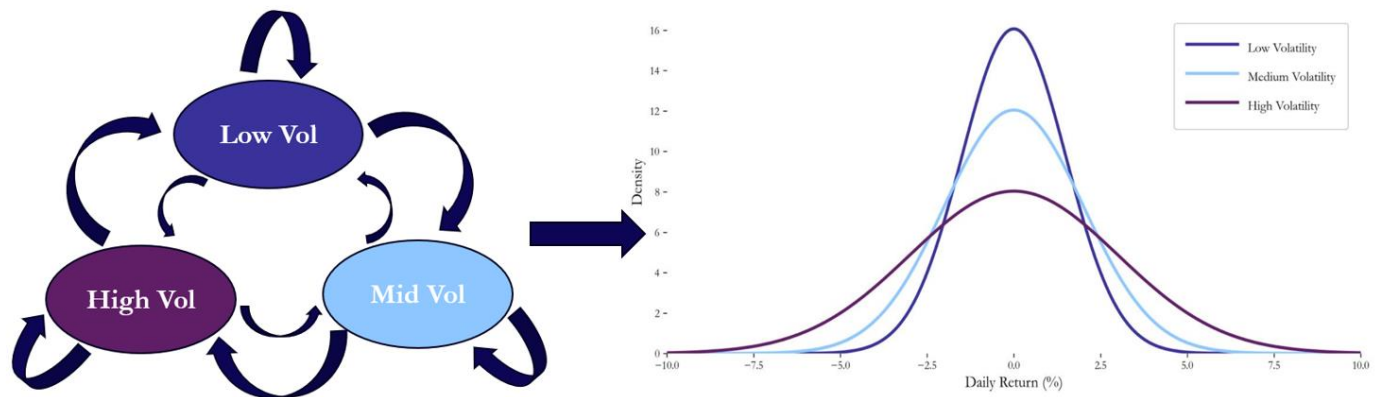
Finally, the third component of the chain is the initial probability distribution $\pi=(\pi_1, \pi_2, \pi_3)$ that specifies the probability of starting in each state.

There are other important assumptions, such as time-homogeneity, meaning that transition probabilities remain constant over time, and a finite state space, which means that the system can occupy only a limited number of states. Other assumptions exist as well, but it is beyond the scope of this article to analyze them.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Example of latent process: Vol

As we saw in the diagram above, we reduced this latent process (volatility) into three vol states based on historic vol and predefined percentile (below the 33rd percentile is classified as *Low Vol*, between the 33rd and 66th percentile as *Mid Vol*, and above the 66th percentile as *High Vol*), estimating from the data. By reducing this latent process into a state transition space, we can produce the estimated state (for a given day or week) that is going to produce the regime distribution that is going to give us better likelihood for probabilities of events.



Source: BSIC

Now we are able to capture dynamics where, during periods of low volatility, we have a different distribution than in periods of high volatility. Of course, volatility is not the only latent process that drives the “true” return distribution. We need to construct a better model that includes the latent processes most relevant for obtaining a better conditional distribution.

For example, we could also consider trend, momentum etc... This leads to a reasonable question: which latent processes are truly relevant? Too many states cause overfitting; too few fail to capture the underlying dynamics.

An important thing to remark is that with Markov chains we are explicitly defining the criteria for the underlying state space. In many cases, however, the events we are interested in are hidden. For this reason, we need Hidden Markov Models as a way of relating a sequence of observations to a sequence of hidden states that explain the observations. With this latter model we don't need to specify anymore the latter processes because the states are learnt directly from the data.

Hidden Markov Models

HMMs are widely used probabilistic models; Jim Simons himself, in his latest book, describes how they played an important role at Renaissance Technologies.

An HMM is characterized by a hidden process that evolves through time according to Markov dynamics, and an observable process that emits data depending on the current hidden state.

Formally, the model is defined by a set of hidden states $Q = \{q_1, q_2, \dots, q_n\}$, a transition probability matrix $A = [a_{ij}]$ where each element a_{ij} represents the probability of moving from state i to state j , an emission matrix $B = [b_i(o_t)]$ that specifies the probability of observing a value o_t given the current state q_i , and an initial probability distribution $\pi = [\pi_1, \pi_2, \dots, \pi_n]$ describing the likelihood of starting in each state.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

A first-order hidden Markov model instantiates two simplifying assumptions: the first one as we said is the “Markov assumption” and second the probability of an output observation depends only on the state that produced the observation q_i and not on any other states or any other observations. This latter assumption is called “Output Independence”.

Once we know the components of the HMM, three fundamental problems arise:

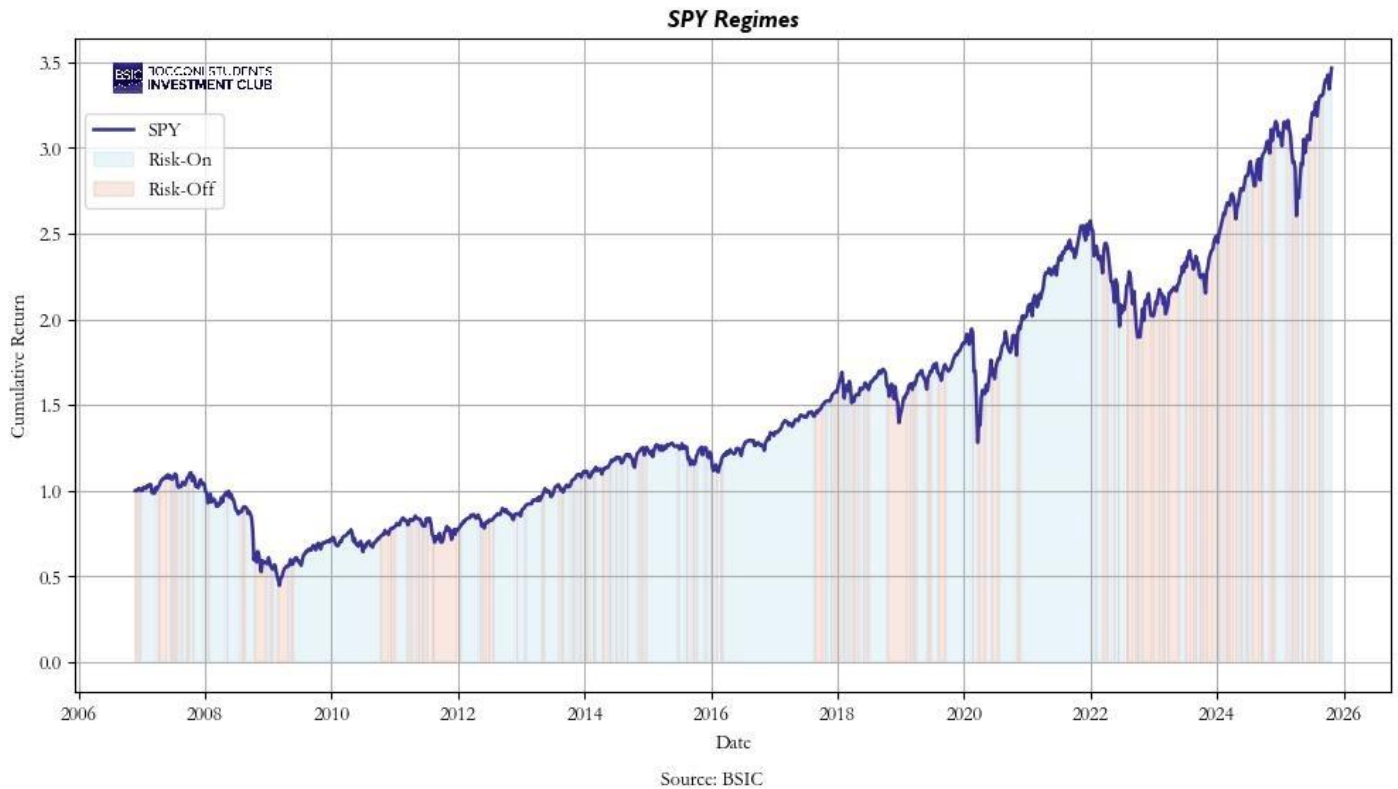
1. Evaluation problem: Given the HMM parameters $\lambda=(A,B,\pi)$ and a sequence of observations $O=(o_1,o_2,\dots,o_T)$, what is the likelihood $P(O|\lambda)$? This is solved by the Forward Algorithm, which efficiently computes the total likelihood of the observed data by summing over all possible hidden state sequences.
2. Decoding problem: Given the HMM parameters $\lambda=(A,B,\pi)$ and a sequence of observations $O=(o_1,o_2,\dots,o_T)$, which sequence of hidden states most likely generated these observations? This is solved by the Viterbi Algorithm, which identifies the most probable path through the hidden states by comparing all possible transitions and keeping, at each step, only the most likely one.
3. Learning problem: How can we estimate the set of parameters $\lambda=(A,B,\pi)$ that makes the observed sequence most likely under the model? This is solved by the Baum–Welch Algorithm (or Forward–Backward Algorithm), which iteratively refines the estimates of A , B , and π . At each step, the model uses the current parameters to estimate how likely it is to be in each state or to move between them, then updates those parameters to increase the overall likelihood of the observed data, repeating the process until the model converges.

Building Models

As a first step we want to identify two regimes based on two different latent processes: SPY returns and volatility. More precisely, to model these regimes, we trained a Gaussian HMM on a rolling window of the data. At each iteration, the model was trained on the most recent observations of SPY weekly returns and the rolling average of the last four weeks of the ETF’s standard deviation, learning the probability distribution and transition dynamics between the two hidden states. In this way we detected two regimes: one characterised by high market returns and

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

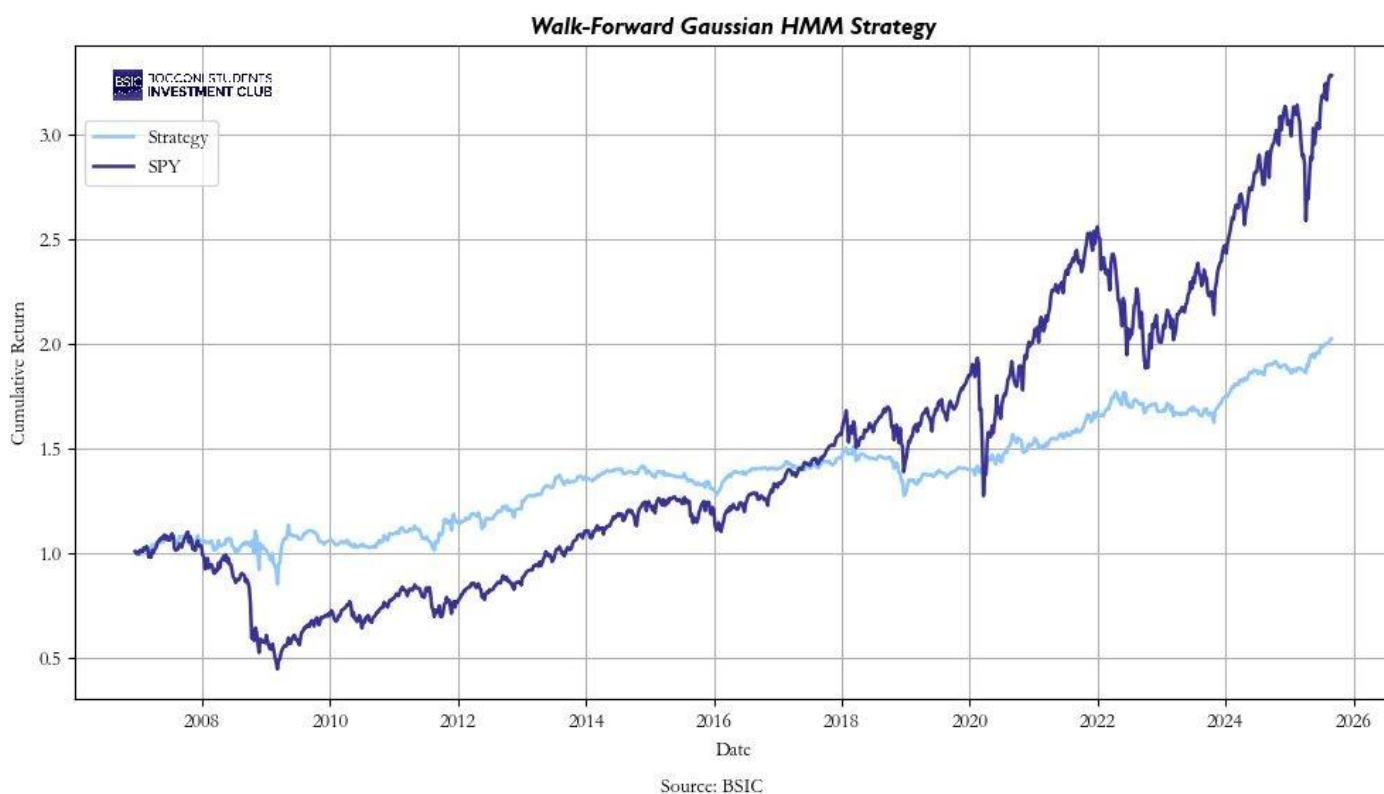
low volatility and the other characterised by negative or low returns and high volatility.



Once the two regimes were identified, we designed a strategy based on regime detection with a switching allocation between the SPY and the value factor (HML). The objective was to gain more exposure to the SPY during risk-on periods while allocating a larger portion of the portfolio to the value factor during distressed or high-volatility regimes.

As before, the model relied on the SPY's weekly returns and the four-week rolling average of its standard deviation as input features and a full covariance matrix was employed to “relax” the assumption of independence between returns and volatility. Using a walk-forward analysis, the HMM was retrained on a rolling two-year window, and for each training window the model produced posterior probabilities of being in each hidden state. At each step, the model used the most recent posterior probabilities and the estimated transition matrix to compute the predicted probabilities for the next period's states. The weight assigned to HML was set equal to the predicted probability of entering the high-volatility state, while the SPY weight was defined as $1 - \text{whml}$. Finally, the portfolio return was computed, rebalanced weekly according to the updated probabilities.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

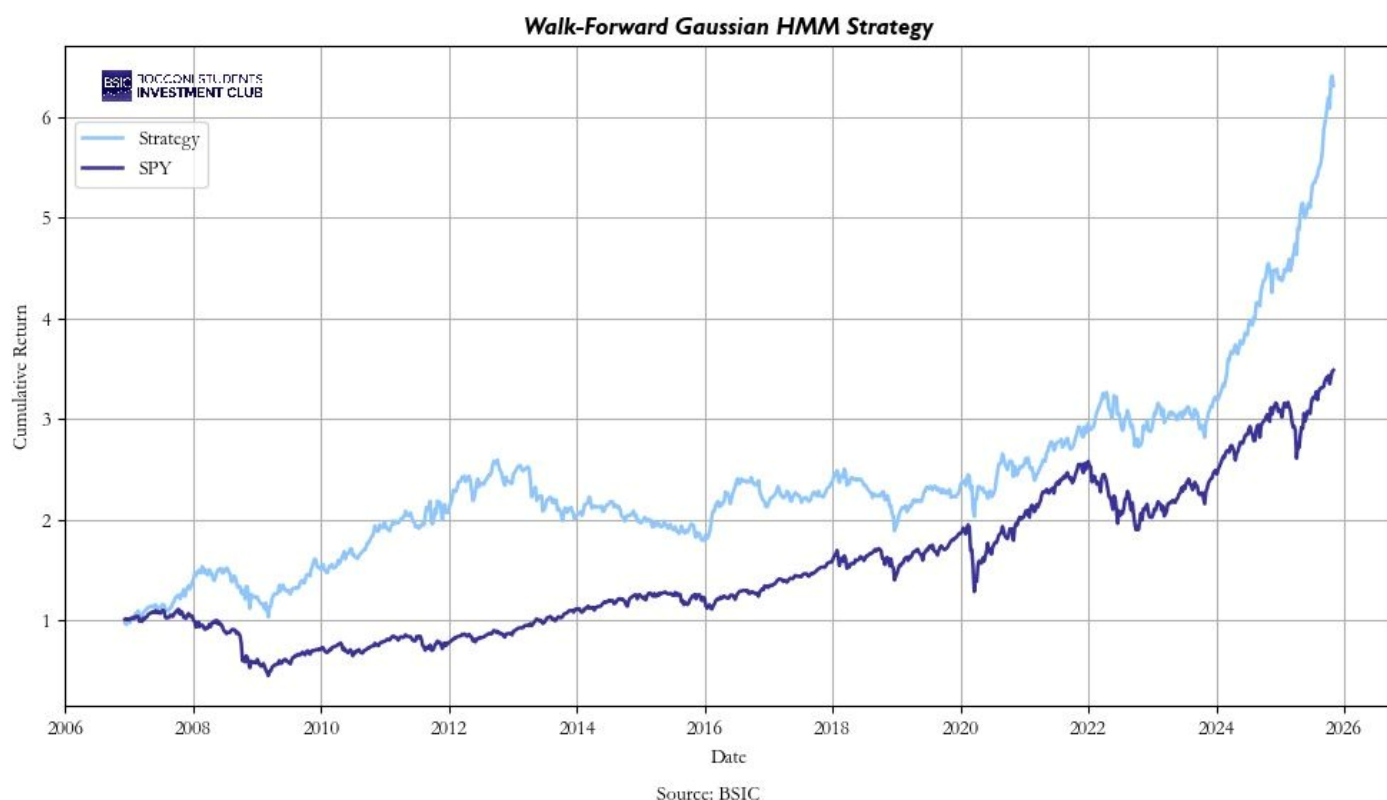


As we can see from the chart, our strategy appears to underperform the benchmark. However, from the table below we observe that this approach consistently reduces volatility and drawdowns, resulting in a higher Sharpe (without accounting for transaction costs).

	CAGR	Max Drawdown	Average Drawdown	Sharpe Ratio
Strategy	5.29%	-23.99%	-5.33%	0.527
SPY	6.53%	-59.56%	-14.25%	0.436
HML	0.48%	-28.59%	-13.75%	0.116

To further develop this approach, we applied the same methodology as before but replaced the “risk-off” variable with the SPDR Gold Shares (GLD).

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.



As we can see, unlike the value factor, gold tends to react more strongly to market stress, providing a more immediate hedge during high-volatility periods thus improving our previous strategy and resulting in a better Sharpe.

	CAGR	Max Drawdown	Average Drawdown	Sharpe Ratio
Strategy	10.19%	-32.69%	-10.52%	0.709
SPY	6.79%	-59.56%	-14.20%	0.450
GLD	8.14%	-48.05%	-22.01%	0.552

References

1. Jurafsky, Daniel & Martin, James H., “Hidden Markov Models”, 2025
2. Wang, Matthew; Lin, Yi-Hong; Mikhelson, Ilya, “Regime-Switching Factor Investing with Hidden Markov Models”, 2020
3. Zucchini, Walter; MacDonald, Iain L.; Langrock, Roland, Hidden Markov Models for Time Series: An Introduction Using R, 2nd ed., 2016.

TAGS: HMMs, Markov Chains, Factors

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.