THE BULLS AND THE BEARS: REGIME IDENTIFICATION AND FORECASTING

M. KING, T. LARSEN, M. SHUMWAY, N. TODD, B. WAITS

Abstract. Financial markets are often classified into two regimes: bull and bear. Identifying the current regime helps investors develop effective strategies. This work uses clustering algorithms, hidden Markov models (HMMs), and a statistical jump model to classify historical regimes and make future predictions upon which an investment strategy is built. While clustering and HMMs produce noisy estimates, the jump model yields smoother regime transitions by penalizing regime switches. Overall, the models effectively classify regimes, and the investment strategy delivers reasonable results.

1. Problem Statement and Motivation

Investing is complicated and often confusing. While financial firms offer many retirement and investment products, the inner workings of these strategies are rarely transparent. Academic investing research can also be difficult to apply in practice. This disconnect motivates our research in exploring mathematical models which may also be applied to real world investing. Specifically, we investigate financial market regime identification and forecasting.

Market regimes are periods where financial markets may be classified differently due to economic conditions or investor sentiment. Recognizing regimes such as high volatility periods versus stable periods or recession phases versus growth phases has been shown to improve investment decision making across asset classes and markets [MSS+12]. Importantly, these regimes appear consistently across various asset classes and markets around the world, making this topic universally relevant.

Prior studies have used various statistical tools to identify these regimes, including HMMs [AT10, HL90] and statistical jump models [NLM20, SYM24]. We build on this by comparing these approaches side-by-side and introducing our own work using the Gaussian mixture model. Our analysis culminates in an investment strategy informed by regime predictions.

In this work, we consider bull and bear financial market regimes. Bull markets characterize periods of large investment gains, low volatility, and positive investor sentiment, whereas bear markets characterize the opposite

Date: April 12, 2025.

behavior. Professional investors are motivated to adjust their portfolio holdings according to this bull and bear model. We therefore seek to construct our investment strategy based on predicted bull or bear market regimes.

2. Data

Our dataset consists of daily price data on 12 exchange traded funds (ETFs): SPY, tracking the S&P 500 index, and 11 sector-based ETFs which together comprise the S&P 500, such as technology, healthcare, and utilities. Our data runs from 1994 to present. This data is sourced using the yfinance package in Python. This package, widely used and carefully maintained, efficiently provides clean and well-organized financial data. While yfinance provides several features in its time series, daily price is the only one relevant to our project. We use price to construct synthetic features.

2.1. **Data Engineering.**

One advantage of using yfinance is that it provides data without missing values or NaNs, largely due to its focus on daily frequency data. This greatly simplifies the cleaning process. As a result, we focus less on cleaning and more on feature construction.

We use daily price data to compute several key metrics, including returns, downside deviation, and Sortino ratios across various time horizons. Daily data strikes a balance between noise and granularity—minute data is too noisy, while weekly or monthly data is too coarse to capture realistic regime shifts. The inclusion of Sortino ratios and downside deviation is motivated by the work in [SYM24]. Returns are calculated as percentage changes in closing price between consecutive days, reflecting profit or loss.

A Sharpe ratio indicates the risk adjusted return or amount of return for one unit of risk with higher values representing stronger performance. The Sortino ratio, a modification of the Sharpe ratio, measures the risk-adjusted return of an asset while specifically considering downside volatility. It is computed by subtracting a target or risk-free return from the asset's actual returns and dividing by the downside deviation. The downside deviation itself is a specialized volatility metric that focuses on negative returns, calculated by zeroing out positive values and computing standard deviation among the resulting values. By calculating downside deviation, investors can better understand the risk associated with undesirable volatility.

We also compute exponentially weighted moving averages (EWMAs) of each feature over several time horizons to capture persistence in features over time. Naturally, the EWMA weighs recent behavior over the past. The length and number of time horizons is a key hyperparameter of our model. For example, one choice we tested was EWMAs with 5, 10, and 60-day half-lives, to simultaneously stress recent and former market conditions.

2.2. **Train-Test Split.**

Identifying stock market regimes using machine learning poses the challenge of ensuring detected patterns remain robust out of sample. To achieve this goal, our research employs a train/test split. This approach is driven by two key motivations:

(1) Assess whether regime classifications derived from historical market data continue to hold predictive power when applied to future (out-of-sample) periods.
(2) Reduce noise and minimize frequent transitions between regimes, leading to clearer and more stable classifications that are less affected by transient market behavior.

We conduct most of our analysis using the SPY dataset, spanning from January 1994 to the present. The data is divided chronologically into a training period (January 1994 – January 2019) and a testing period (January 2019 onward) to prevent temporal leakage. The cutoff in 2019 was chosen as a splitting point because we wanted to include the COVID-19 market crash and rebound, and the 2022 bear market. The models are exclusively fitted on the training data and then evaluated on the testing data.

3. Methods

Regime identification can be viewed both as a clustering problem, where each market regime corresponds to a different cluster, and as a latent variable problem, where hidden states are market regimes and observable states are financial metrics like prices and returns. As investment regimes typically persist for several weeks or months rather than days, model behavior should reflect this economic intuition by exhibiting a high degree of persistence, characterized by few regime switches.

This project explores various methods to classify the regime of the market, including K-Means clustering, Gaussian mixture model (GMM), Gaussian hidden Markov model (GHMM), Gaussian mixture model hidden Markov model (GMMHMM), and a statistical jump model. Each of these are explained in greater detail in the following.

3.1. K-Means.

As an initial approach, we applied the K-Means algorithm to partition market conditions based on historical data. This is a hard clustering algorithm, meaning that each data point is assigned to exactly one cluster with no uncertainty or overlap. With two clusters, the model effectively distinguished between bull and bear markets; however, it resulted in frequent switches between regimes over time, which is inconsistent with the expected persistence of market conditions.

3.2. Gaussian Mixture Model.

The GMM, of which K-Means is a specific case, was explored as the second model. K-Means implicitly assumes that clusters are symmetric and equally sized, which can be overly simplistic for modeling the structure of investment pricing data. This model assumes that data originate from a mixture of Gaussian distributions, allowing market data to be clustered by learning both the mean and variance for each regime. Using a SciKit-Learn GMM fitted to two components, the clustering results were similar to those of K-Means in that regimes were effectively identified. For GMM, the results demonstrated even more frequent regime switches than K-Means, likely attributed to GMM dynamically adjusting covariance and providing a more precise fit. A grid search was conducted across parameters such as covariance type and number of components to refine the model. However, the results found limited benefit of parameter adjustments in reducing regime switches, which validated the use of the out-of-the box model.

3.3. **Gaussian Hidden Markov Model.** The HMM framework is suitable for regime identification for time-indexed financial data as it predicts states sequentially, with previous dates influencing the next. The Gaussian hidden Markov model (GHMM) is an HMM which assumes that emission probabilities (i.e. the distribution from which observations are generated from hidden states) follow a Gaussian distribution. In the case of regime identification, the space of hidden states is {bear, bull} and the space of observations are daily returns and engineered features, as described above. The GHMM class from the HMMLearn Python package was used for all experiments.

3.4. **Gaussian Mixture Model Hidden Markov Model.** Another approach used in regime prediction modeling is the Gaussian mixture model hidden Markov model (GMMHMM). This differs from the GHMM in that emission probabilities are now modeled with a Gaussian mixture model rather than a single Gaussian.

In our implementation, we used the HMMLearn Python package to efficiently apply the GMMHMM model using two components. Similar to the GMM, a grid search was conducted on the model, with negligible benefit. Both HMM methods, like K-Means and GMM, successfully identified regimes but displayed large amounts of noise, switching bull and bear identifications faster than real-world markets.

3.5. **Statistical Jump Model.** A promising model that mitigates unrealistic switching patterns is the statistical jump model, first introduced in [NLM20]. The objective of this method is to fit a regime identification model which penalizes switching between regimes. The formal definition given below is adapted from [AKMS24]. Specifically, given an observation sequence $\mathbf{Y} := \{\mathbf{y}_0, \ldots, \mathbf{y}_{T-1}\}$ with $\mathbf{y}_t \in \mathbb{R}^D$ for all t, a statistical jump model with K states is given by solving the optimization problem

$$\underset{\Theta,S}{\arg\min} \sum_{t=0}^{T-1} \frac{1}{2} \|\mathbf{y}_t - \theta_{s_t}\|_2^2 + \lambda \sum_{t=1}^{T-1} \mathbb{1}_{s_{t-1} \neq s_t} \tag{1}$$

where $\Theta := \{\theta_k \in \mathbb{R}^D : k = 0, \ldots, K-1\}$ are the model parameters, $S := \{s_0, \ldots, s_{T-1}\}$ denotes the state sequence, and $\lambda \in \mathbb{R}_+$ is a hyperparameter referred to as the jump penalty. A larger value of λ encourages regime persistence, where a smaller value allows more flexibility. We can interpret the loss function as a balance between fitting the data with multiple models and prior beliefs on the persistence of the state sequence.

For implementation purposes, we used the python library jumpmodels that was created along with the work in [AKMS24]. Notably, this library learns the parameters of the jump model by performing a coordinate descent algorithm several times, each run initialized by the results of the K-Means algorithm. The parameters that result in the lowest value in the objective function are then kept. For our modeling, we chose $\lambda = 50$, as suggested by the authors in [AKMS24]

The empirical results from the jump model demonstrate effective regime identification while maintaining a reasonable rate of state switching. A comparison between the test set performance of the statistical jump model and the four previous models is given in Figure 1.

3.6. **Dimension Reduction Techniques.** A key challenge in stock market regime identification is ensuring that regimes are both robust and practically useful. Frequent switching between regimes—often driven by short-term market noise—can raise transaction costs and reduce interpretability. To address this issue, we applied Principal Component Analysis (PCA) to reduce data dimensionality and suppress noise. By capturing most of the variance in a smaller number of components, PCA can increase the signal-to-noise ratio, potentially resulting in fewer and more meaningful regime transitions.

Training features were standardized, and components explaining 95% of cumulative variance were retained. To prevent data leakage, the PCA transformation was derived from the training set and directly applied to the test set. Effectiveness was assessed by visualizing cumulative SPY log-returns with color-coded regimes, evaluating stability both qualitatively (continuity) and quantitatively (switching frequency).

Despite its theoretical appeal, PCA yielded only modest practical benefits. It slightly reduced regime switching in the KMeans model but showed limited improvements in overall stability or interpretability across other methods. Experimenting with random Gaussian projections yielded similar results so we will omit it's discussion here. In this context, dimension reduction techniques contributed little to reducing noise, reinforcing the conclusion that the observed instability is not primarily due to high dimensionality.

3.7. **Forecasting.** Forecasting stock market regimes remains a notoriously challenging yet essential task for investment management. Accurate regime predictions enable investors to strategically adjust portfolios ahead of major market shifts, potentially enhancing returns and limiting risk more effectively. Fortunately, forecasting investment regimes is simpler than predicting returns directly. Investigating bull and bear markets, we are only concerned about forecasting a binary indicator of up or down rather than the specific magnitude of returns in continuous space.

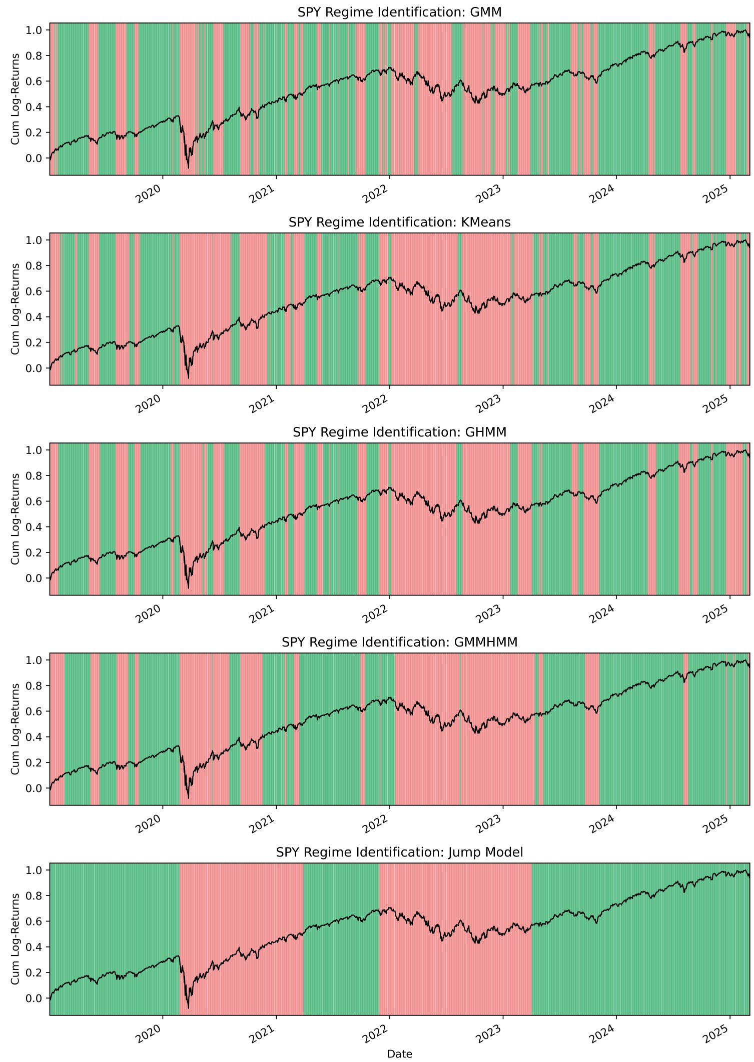

FIGURE 1. Regime identification on test set for several models fit on the training data. Each model partitions the data into roughly the same regimes, however, all except the statistical jump model have a high frequency of switching. In this figure, green corresponds to a bull identification and red corresponds to bear.

In order for forecasts to be applied directly to investment strategies, predictions must be made in an online or rolling fashion, contrary to batch prediction utilized in the test/train split mentioned previously. To construct forecasts for the S&P 500, we fit each of the five models previously mentioned with a 10 year lookback for each day, t. We then used built-in

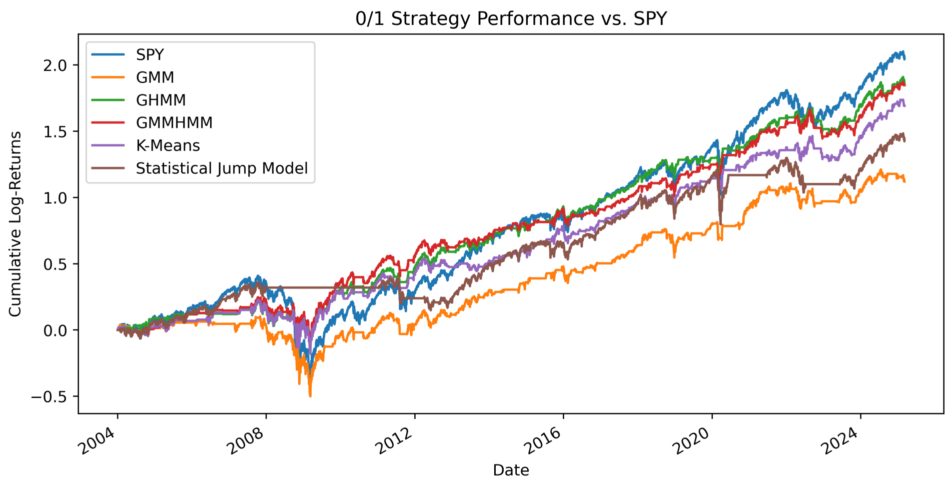

FIGURE 2. Plot of the returns of the S&P 500 Index (SPY) against each of the forecasted model results. We see that the statistical jump model excelled at avoiding the great recession market downturn, and the other four methods limited losses during the COVID-19 market crash. Note that although the S&P has higher returns, the Sharpe Ratio of several models is higher (see Table 1).

functions (.fit()) to estimate the most likely regime sequence corresponding to the data of days $t - 10$ years to t. The last estimated value for day t was then rolled forward one day to act as the prediction for day $t + 1$. This method relies upon a stationarity assumption between consecutive days, which is realistic given economic intuition that bear and bull markets persist for long durations. Because our data begins on January 1, 1994, the first forecast occurs at the beginning of 2004. To further enhance regime persistence, forecasts were filtered using a 12-day rolling median.

Because these forecasts were generated without look-ahead bias, trading strategies can be created. We employed a simple 0/1 investment strategy which deploys 100% capital to SPY if the forecasted regime indicates bull or 0% capital if bear is indicated. The performance of the strategy is measured by the Sharpe Ratio. Figure 2 displays the performance of each model against buying and holding SPY, comparable to Exhibit 9 shown in [SM24].

The jump models package published in tandem with [SM24] includes an online prediction method which generates a temporal state sequence day by day, enabling improved application to investment strategies. We fit statistical jump models for each of the 11 sector ETFs with daily data from 2005 - 2019. The online prediction method then predicted future investment regimes to present time. Because sectors consist of fundamentally different companies, the sector ETFs are diversified from one another, with sector specific bear and bull markets occurring asynchronously with other sectors. This phenomenon enables dynamic sector rotation, which we simulate by investing equal amounts of capital into each sector indicating bull market for each day t. The performance can be visualized in Figure 3. It should be noted that while the dynamic sector strategy achieves a higher Sharpe ratio than SPY and mitigates some crash risk, there still exist some days where all but one ETF signals bear market, leading to significant losses when the one bull prediction is incorrect on a market crash day. As such, future research should investigate portfolio optimization techniques which consider forecasting error.

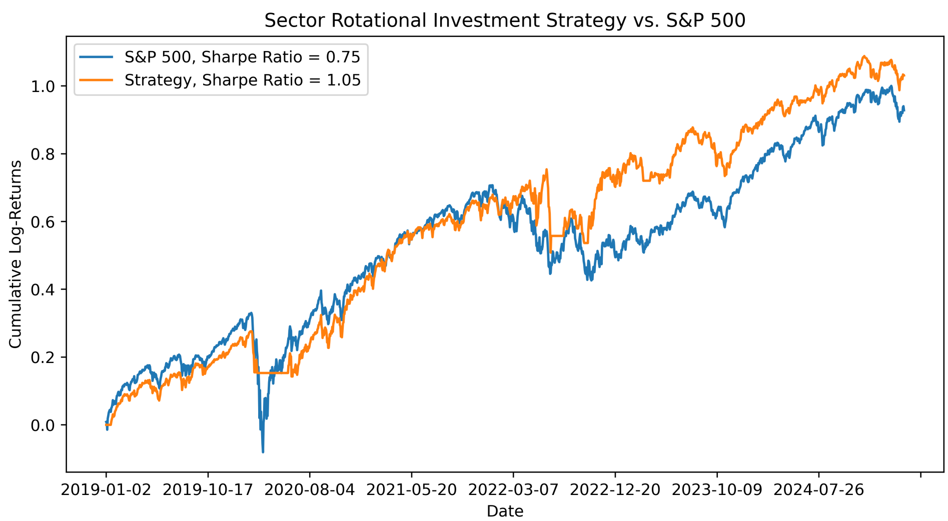

FIGURE 3. The dynamic sector rotational investment strategy outperforms the S&P 500 on a risk-adjusted basis. The strategy effectively mitigated part of the COVID-19 market crash, yet it still suffered large drawdowns at other times.

4. RESULTS AND ANALYSIS

The models evaluated in this paper demonstrated varying levels of effectiveness in identifying market regimes. Two key performance measures were used: visual identification of general market trends (to ensure correct detection of major bull and bear periods) and average yearly regime changes, which captures model confidence and stability. These metrics together offer a balance between interpretability and practical applicability for trading strategies. Based on this evaluation, Table 1 provides summary statistics for our custom online forecasting method as described in section 3.7. Figure 1 qualitatively displays similar results for our test/train split modeling.

As shown in Table 1, the statistical jump model performed best, correctly aligning with market trends and averaging only 0.70 regime changes per year. The GMM also identified key bull and bear phases, but with a higher average of 8.80 yearly regime changes, indicating some noise and less stability. In contrast, K-Means, GHMM, and GMMHMM exceeded 12.00 regime changes per year, making them less practical for trading. Although these models were generally successful in detecting market downturns, the frequency of regime shifts reduced their reliability, and none showed a clear advantage over the others.

	GMM	**K-Means**	**GHMM**	**GMMHMM**	**JM**
Mean Yearly Regime Changes	8.80	14.80	12.80	13.60	0.70
Mean Regime Length	51.70	31.70	33.30	32.80	380.60
0/1 Strategy Sharpe Ratio vs. SPY (0.51)	0.38	0.59	0.65	0.64	0.50

Table 1. Summary statistics of forecasting models using a 12-day rolling median. The jump model shows a significantly longer average regime length, suggesting it is best at limiting noise in market data.

We believe the primary reason the statistical jump model outperformed the others is its explicit incorporation of a penalty for regime switches. This is further supported by the fact that setting the penalty term $\lambda = 0$ in the objective function (Equation 1) reduces it to the K-Means objective—corresponding to our K-Means model, which produced noticeably noisier predictions. The remaining models (GMM, GHMM, GMMHMM) are all built on the assumption of normally distributed data. However, since these are probabilistic models without any explicit penalty for regime switching, there is little reason to expect their estimates to be smooth or resistant to noise in the absence of additional regularization. A variant on these probabilistic methods wherein regime switches are penalized would be an interesting future research direction.

Our work with dimensionality reduction techniques further suggests that the artificially high dimensionality of the dataset does not significantly contribute to the noise within the models. Rather, the instability appears to stem from inherent market complexity and the limitations of the models themselves, rather than from excess input dimensions.

Finally, our forecasting models and trading strategy outperformed the buy-and-hold S&P 500 approach, achieving a higher Sharpe ratio while mitigating market crash risk. This demonstrates the potential of regime forecasting despite the inherent challenges of financial market prediction. While extensive research exists in this area, our results add GMM models to the current body of literature and reinforces the value of statistical models in predicting and classifying regimes, which can be leveraged to enhance portfolio trading strategies.

5. Ethical Implications and Conclusions

In financial markets, ethical risks are inherent in the handling of money, especially when making predictions that influence investment decisions. Although the data used in our models is publicly available via the yfinance package, the predictive nature of these models still raises important ethical considerations. While privacy is not a concern in this case, the broader responsibility of using predictions to guide financial behavior remains. Firms offering financial advice typically undergo certification in risk management and ethics to ensure legal and ethical compliance. A key concern is black-box decision making, where models classify market regimes without clear, interpretable rationale. This lack of transparency can undermine trust, accountability, and justification for decisions based on such predictions, potentially leading to misguided investments and financial harm.

We also recognize there exists a potential risk of our model creating a feedback loop. For example, if the model predicts a bull market, investors might act on the prediction and help create the trend that was predicted. However, we believe the market's behavior is not purely driven by speculation. Prices are still connected to the real value of businesses, which generate revenue and hold assets. So while market sentiment matters, our work aims to capture deeper, meaningful shifts based on actual economic activity.

6. Conclusion

In this study, we evaluated several machine learning tools for financial regime identification and forecasting using ETF data from 1994 onward. Given the results, it is clear that the statistical jump model should be prioritized over other methods when developing a regime-based investment strategy. Our PCA-based dimensionality reduction approach yielded marginal improvements but did not significantly enhance model stability. We recognize the limited applicability of our modeling to the average investor because of computational requirements. For the investment professional, while quantitative regime identification techniques hold promise, significant challenges remain before such methods can reliably guide investment strategies without additional risk management safeguards. We propose that future work could consider incorporating penalties for regime switches into the GMM, GHMM, and GMMHMM models.

References

[AKMS24] Afşar Onat Aydınhan, Petter N Kolm, John M Mulvey, and Yizhan Shu. Identifying patterns in financial markets: extending the statistical jump model for regime identification. *Annals of Operations Research*, pages 1–37, 2024.

[AT10] Andrew Ang and Allan Timmermann. Regimes and the cross-section of returns, 2010. MPRA Paper No. 21154.

[HL90] James D. Hamilton and Gang Lin. Stylized facts of daily return series and the hidden markov model. *Journal of Applied Econometrics*, 5(1):69–93, 1990.

[MSS+12] Michael C. Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H. Seligman, Thomas Guhr, and H. Eugene Stanley. Identifying states of a financial market. *Scientific Reports*, 2:644, 2012.

[NLM20] Peter Nystrup, Erik Lindström, and Henrik Madsen. Learning hidden Markov models with persistent states by penalizing jumps. *Expert Systems with Applications*, 150:113307, 2020.

[SM24] Yizhan Shu and John M Mulvey. Dynamic factor allocation leveraging regime-switching signals. *arXiv preprint arXiv:2410.14841*, 2024.

[SYM24] Yizhan Shu, Chenyu Yu, and John M Mulvey. Downside risk reduction using regime-switching signals: A statistical jump model approach. *Journal of Asset Management*, 25(5):493–507, 2024.