Using Large Language Models (LLMs) for Investment

Elina (Fuwei) Zhuang: fz266
Hudson Chen: hc884
Shun Wang: sw2337

Summary

Researchers often find that the performance of strategies can vary significantly depending on market conditions. Intuitively, regime-switching strategies may offer more adaptive and consistent returns. In this paper, we aimed to identify different market regimes based on both market performance and investor sentiment by leveraging the Gaussian Mixture Model (GMM) and Large Language Models (LLMs). We then calculated the probability of the next day's regime and used these probability distributions to compute the mean and covariance matrix. Applying Modern Portfolio Theory (MPT) to these results, we determined allocations, ultimately outperforming the S&P 500.

Compared to usual regime switch strategies, our **2D-Bayesian Regime Switch Strategy** had three major innovations:

- **2-Dimensional regimes**: instead of just keeping track of recent market trends using GMM, our approach also incorporated market sentiments leveraging LLM's edge (GPT-4o) of interpreting market news, which was a significant driver in market movements. By employing the product of vector space, the strategy should sufficiently capture the interaction between these two regimes.
- **Rational and adaptive decisions**: Instead of implementing strategies according to the most likely regime, we tailor decisions based on expected returns and covariance matrix according to regime probability distributions. For example, the strategy should be differentiated when a "neutral" condition will happen with a probability of 51% and 100%. By including Bayesian updates, the strategy is constantly learned from markets. So in the long run, the performance was better.
- **Proper fit to MPT:** MPT assumes the returns are normally distributed with constant correlations. After the implementation of GMM, each regime's return was specifically fitted into normal distributions. Within the regime, the correlation coefficients were relatively steady compared to pure returns. Thus, MPT had a better performance here than usual.

Methodology

Step 1: Market Regime Labeling Based on Historical Return Data using Statistical Model - GMM

In the context of regime detection, we model the returns of a set of factors as a mixture of *K* Gaussian distributions. Each Gaussian distribution represents a regime.
The factor returns are denoted by $f_t = (f_{1t}, f_{2t}, \ldots, f_{Nt})'$, where f_{it} is the return of the *i-th* factor at time *t*.
The factor returns are assumed to be generated by the following mixture model:

$$p(\mathbf{f}_t) = \sum_{k=1}^{K} \pi_k \phi(\mathbf{f}_t; \mu_k, \Sigma_k)$$

where π_k is the mixing proportion of the *k-th* regime, $\varphi(f_t; \mu_k, \Sigma_k)$ is the density of a multivariate Gaussian distribution with mean μ_k and covariance matrix Σ_k . We use the Expectation-Maximization (EM) algorithm to estimate the parameters of the model and assign each observation to a regime.
Our choices of factors include 9 sector ETFs to indicate the performance, the Fama-French 5 factors, the inflation rate, the interest rate, and the momentum factor. Details of the factors are shown in Table 1. To avoid overfitting, we use the principal components analysis to reduce the dimension of the factor returns to 3, and use the Bayesian Information Criterion (BIC) to select the number of regimes, which is 5 in our case.

Table 1: Details of factors

Factor	Description
XLF	Financial Select Sector SPDR Fund
XLV	Health Care Select Sector SPDR Fund
XLY	Consumer Discretionary Select Sector SPDR Fund
XLI	Industrial Select Sector SPDR Fund
XLB	Materials Select Sector SPDR Fund
XLE	Energy Select Sector SPDR Fund
XLU	Utilities Select Sector SPDR Fund
XLK	Technology Select Sector SPDR Fund
XLC	Communication Services Select Sector SPDR Fund
MKT	Market return
SMB	Small Minus Big
HML	High Minus Low
RMW	Robust Minus Weak
CMA	Conservative Minus Aggressive
INFL	Inflation rate
RF	Interest rate
MOM	Momentum factor

Step 2: Market Regime Labeling Based on Investor Sentiments Using Large Language Models (LLMs)

We utilized large language models from January 1, 2012, to December 31, 2019, using a Retrieval-Augmented Generation (RAG) framework with GPT-4o for market regime summarization. For predictions, GPT-4o mini was employed from January 1, 2012, to July 31, 2024, with out-of-sample testing from 2020 to 2024. The Wall Street Journal provided the only comprehensive dataset suitable for U.S. equity market analysis. Using the RAG framework and Llama-Index, we summarized daily news into structured datasets, stored in a Milvus vector database for efficient retrieval. After prompt engineering, the model classified market regimes into three categories (Negative, Neutral, Positive) and performed daily predictions using refined prompts and sampling parameter tuning for quality outputs.

Step 3: Estimating the probability distribution of the Next-Day market regime (from Step 1 & 2) - Bayesian Updates

First of all, we have assumed the statistical regimes follow a multivariate Bernoulli distribution. In simple words, for regime i, it has a probability of P_i, where P_i's add up to 1. We considered it make most sense to use Bayesian Updates for the following reasons:

1. **Probability Distribution**: instead of returning a most likely regime, it should give probabilities of all regimes. Since some regimes (e.g. neutral) happened with a high frequency, it was likely to predict those regimes, and ignore the rare ones (e.g. Great Depression.) Thus, using Bayesian Prior to give probabilities was more optimal than using complex models to classify.
2. **Behavioral finance**: Differ from academic assumptions, returns are not independent with constant means and standard deviations. This is evidence of the existence of investors' Anchoring Bias. To exploit this bias, Bayesian priors imitated the behavior of investors making decisions based on historical data.
3. **Dynamic market**: since the market is hardly static, a dynamic model outperforms in the long run. The process of updating priors not only caught market trends, but also had a fast running time, which allowed it to trade just before market close without much price slippage.

We used Bayesian rules to model tomorrow's regime based on today's. In more specific:

$$P(A \mid B \cap C) = \frac{P(B \cap C \mid A) \times P(A)}{P(B \cap C)}$$

Where:
A = Tomorrow's Statistical Market Regime,
B = Today's Statistical Market Regime,
C = Today's Sentiment Regime

Note that we also utilized Laplace Smoothing to avoid overseeing rare cases.

We stored the prior as a 3-dimensional dataset. Every day we update the prior, then extract the vector corresponding to today's statistical and sentiment regime. Based on our probability inference, we could make trading decisions.

Step 4: Bayesian-Informed Approach to Dynamic Portfolio Allocation

We apply a Bayesian framework to enhance the classical Mean-Variance Optimization (MVO) framework by incorporating information about different market regimes. This allows us to create a portfolio that is adaptable to changing market conditions. The model is built using the following steps:

1. For each regime identified by GMM, we build an optimal allocation for each market condition by applying the classical MVO framework. This allocation relies on the historical daily returns and covariance matrix specific to that regime. We aim to maximize the Sharpe ratio in each regime for 300 individual stocks in S&P 500 embedded in 9 different sectors.
2. Next, we employ Bayesian updating to dynamically estimate the probabilities of being in each market regime at any given time, allowing the model to incorporate real-time information about market conditions.
3. We then combine optimal portfolio weights with the regime probabilities to construct a weighted average portfolio for each day.

4.	The final step is to dynamically rebalance the portfolio by adjusting the portfolio’s exposure to each of the 9 sectors based on the daily regime probabilities.
This investment strategy is flexible and is capable of responding to changes in market conditions, thus creating a portfolio that is robust under volatile market conditions.

Data Insights

Our GMM outputs 5 conditions based on daily returns. We aggregate the results to show the most frequent regime label for each month from 1999 to 2024.

Market Regimes: Figure 1 illustrates how market conditions evolved over time, with dominant regimes shown as longer stretches of the same color and transitions reflected in shorter intervals with varying colors.

1999 - 2007 (Top): A relatively stable period with brief regime transitions, including volatility during the early 2000s dot-com bubble.

2008 - 2015 (Middle): The most unstable period begins in 2008, with frequent regime shifts corresponding to the global financial crisis.

2016 - 2024 (Bottom): Regime shifts intensify around 2020, aligning with the pandemic and its aftermath, including high inflation.

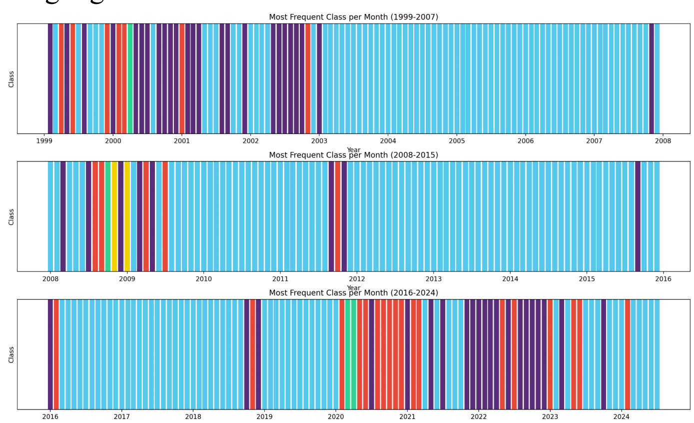

Fig. 1: GMM Market Regime Classifications

LLM outputs are based on the available news dates, with these restrictions LLM provides predictions from 2012 onwards for the labels. Figure 2 shows the distribution of the generated label.

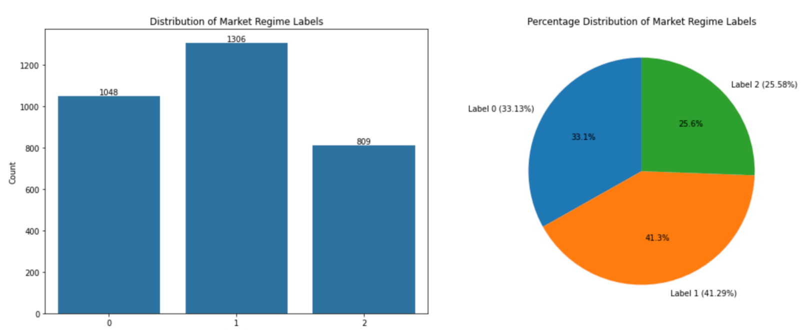

Fig. 2: LLM Market Regime Classifications

Conclusion

With a backtesting period of 4 years (2016-01-05 to 2019-12-31), our strategy consistently outperformed the S&P 500, beating the benchmark by 17.38% in cumulative returns (see Figure 3). The Sharpe ratio increased by 0.09, indicating improved risk-adjusted returns without leveraging risk. As shown in the figure, the strategy maintained consistent excess returns across business cycles, validating the effectiveness of regimes and Bayesian updates.

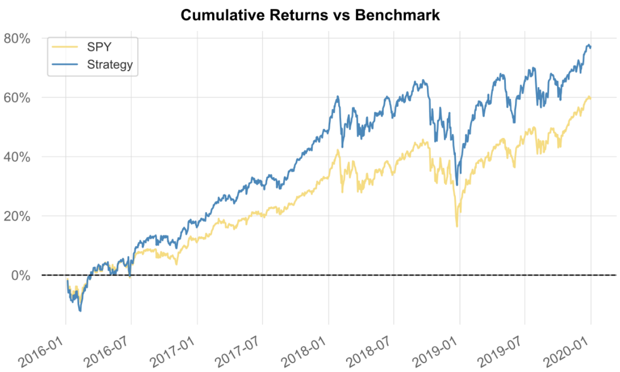

Fig. 3: Cumulative Returns vs. Benchmark (S&P 500)

Key Takeaways:

- **RAG and Database Efficiency**: Milvus effectively stores embeddings but requires careful budget management due to high embedding costs. Checkpointing helps mitigate these expenses.
- **Parameter Tuning and Chunking**: Optimal results depend on careful tuning and chunking data for efficient embedding and retrieval.
- **GMM and MPT Synergy**: Clustering with GMM normalizes returns within regimes, enhancing MPT performance.
- **Bayesian Updates**: Captures anchoring bias, improving adaptability to investor behavior.

Recommendations:

- **Expand Data Sources**: Incorporate additional news sources beyond The Wall Street Journal to increase data diversity.
- **Enhance LLM Stability**: Address instability caused by context length limitations and misclassifications with supervised fine-tuning and post-training on domain-specific data.
- **Optimize Strategy**: Expand to long-short strategies to better adapt to market dynamics and include transaction costs for more realistic performance assessments.

By combining market performance data with investor sentiment analysis, our regime-switching strategy demonstrated the potential for robust, adaptive investment outcomes, outperforming the S&P 500 over the long term.