1 Regime Detection Using Hidden Markov Models

1.1 Objective

The objective of the regime detection module is to identify latent economic regimes for each macroeconomic indicator using historical time-series data. These regimes represent distinct economic states that influence the cross-sectional performance of equity factors. Hidden Markov Models (HMMs) are employed as they explicitly model unobserved states, capture temporal dependence, and allow probabilistic regime inference.

1.2 Macroeconomic Indicators

Each macroeconomic indicator is modeled independently using a univariate HMM. The indicators and corresponding transformations are summarized in Table 1.

Table 1: Macroeconomic Indicators and Transformations

Indicator	Transformation
Inflation	Year-over-year change
Production	Month-over-month change
Sentiment	Level
Market (MSCI World Index)	Month-over-month return
Debt	Month-over-month change
Inflation Expectation	Year-over-year change

The independence assumption across indicators limits model complexity and avoids over-parameterization while retaining sufficient economic interpretability.

1.3 Data Transformation

To ensure stationarity and scale invariance, macroeconomic series are transformed into returns or growth rates according to:

$$R_t = \frac{O_t - O_{t-i}}{O_{t-i}}, \tag{1}$$

where O_t denotes the observed value at time t, and $i = 1$ or 12 corresponds to monthly or yearly changes, respectively.

1.4 Model Specification

1.4.1 Hidden States

A two-state Hidden Markov Model is specified for each indicator. The two regimes are interpreted as:

- **Regime 1**: Low-return or adverse economic regime,
- **Regime 2**: High-return or favorable economic regime.

This binary specification balances interpretability with statistical robustness given the limited length and frequency of macroeconomic time series.

1.4.2 Observation Equation

Conditional on the latent regime q_t, the observed return R_t follows a Gaussian distribution:

$$R_t \mid q_t = i \sim \mathcal{N}(\mu_i, \sigma_i^2), \tag{2}$$

where μ_i and σ_i denote the mean and standard deviation associated with regime i.

1.4.3 Model Parameters

The complete parameter set of the HMM is given by:

$$\lambda = \{A, \mu, \sigma, p\}, \tag{3}$$

where A is the 2×2 transition probability matrix, $\mu = (\mu_1, \mu_2)$, $\sigma = (\sigma_1, \sigma_2)$, and p is the vector of initial regime probabilities.

1.5 Parameter Estimation

At the end of each month, the HMM is re-estimated using all available data up to that point. Model parameters are obtained via maximum likelihood estimation using the Baum–Welch algorithm. This rolling re-estimation framework allows the model to adapt to structural economic changes while avoiding look-ahead bias.

1.6 Regime Identification

Since HMM states are unlabeled, a post-estimation identification rule is applied to ensure consistent economic interpretation. Each regime is ranked according to its risk-adjusted return:

$$\text{Score}_i = \frac{\mu_i}{\sigma_i}. \tag{4}$$

The regime with the lower score is classified as Regime 1, while the regime with the higher score is classified as Regime 2.

1.7 Regime Inference

Following parameter estimation, posterior regime probabilities are computed and the most likely regime at time t is selected. This process yields a binary regime label for each indicator at each time period.

The regime detection module produces a six-dimensional regime state vector at every time t, which serves as input for the subsequent factor allocation and stock selection stages.

1.8 Discussion

The regime detection framework is designed to be parsimonious, interpretable, and robust. The use of univariate HMMs and Gaussian emissions ensures tractability, while rolling estimation captures evolving macroeconomic dynamics. Although the approach abstracts from joint regime interactions across indicators, it provides a stable foundation for conditional factor-based investment strategies.

2 Practical Considerations and Data Constraints

2.1 Use of Correlation Matrix for Indicator Selection

Prior to regime modeling, a correlation matrix was computed across all candidate macroeconomic and alternate data series to guide indicator selection.

This analysis was used to:

- Identify highly correlated variables providing overlapping information,

- Reduce redundancy across alternate data headings,
- Improve interpretability and numerical stability of the regime detection framework.

Indicators exhibiting strong pairwise correlations were not jointly included, ensuring that each Hidden Markov Model captures distinct economic dynamics rather than correlated noise.

2.2 Handling of Yearly and Annual Data

Certain macroeconomic and alternate indicators were available only at an annual or yearly frequency, while the regime detection framework operates on a monthly time grid.

To address this mismatch, annual values were held constant across all months within the corresponding year.

Rationale

Alternative approaches such as linear or spline interpolation were evaluated but resulted in:

- Artificial high-frequency spikes,
- Inflated volatility estimates,
- Unstable HMM parameter estimation and poor model convergence.

Maintaining a constant annual value preserves the true information frequency of the data and prevents the introduction of synthetic dynamics that distort regime identification. This approach produced more stable regime sequences and improved training performance, particularly for indicators capturing slow-moving structural trends.

2.3 HMM Convergence Issues Due to Data Availability

During implementation, instances were observed where the Hidden Markov Model failed to converge, primarily due to limited or sparse data availability.

Convergence issues arose when:

- The effective sample size was insufficient,

- Long gaps or irregular reporting intervals were present,
- Variance collapsed due to repeated constant observations.

Hidden Markov Models require adequate temporal variation to reliably estimate transition probabilities and emission distributions. In cases where convergence was not achieved, indicators were either stabilized using the annual-constant approach or excluded from regime modeling.

Summary of Implementation Trade-offs

Issue	Decision	Justification
Correlated indicators	Correlation-based filtering	Avoid redundancy and instability
Annual / yearly data	Constant within year	Prevent artificial volatility
Sparse data	Stabilization or exclusion	Ensure HMM convergence