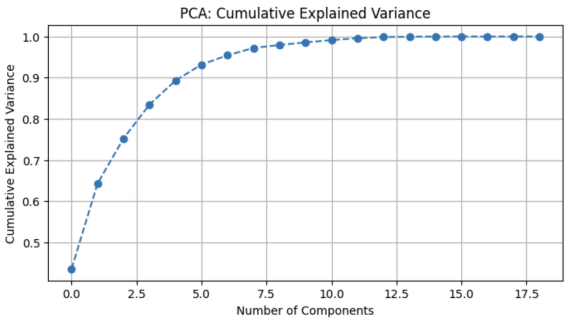
Common market insight is that high positive correlation between major market assets such as stocks, bonds, commodities, and currencies indicates systematic risk and possible market crash. The goal of this project is to find the statistical “edge” to characterize various regimes of the stock market, when assets move together, to see if that has predictive power for future volatility or market crashes. Using a Hidden Markov Model on 19 features engineered from the equity sector and macroeconomic data, the analysis uncovered three distinct market regimes.

This project hopes to answer three questions: Can one detect latent market regimes by clustering the time-varying correlation structure of equity sectors? Do transitions into a "high-correlation" regime predict subsequent volatility spikes or market drawdowns? Does this HMM model beat just watching VIX or simple correlation metrics?

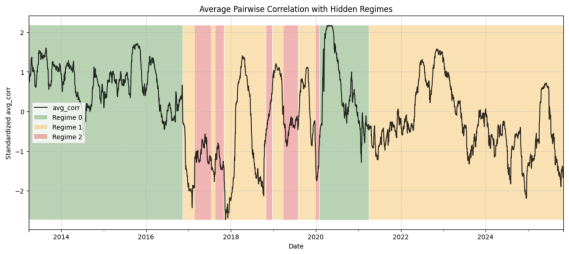
The analysis was conducted in four stages using daily data from 2013 to the present:

1. **Data Collection:** Data was pulled from from 11 key equity ETFs (SPY", "XLK", "XLF", "XLE", "XLY", "XLP", "XLI", "XLV", "XLU", "XLB", "IWM") and macroeconomic indicators ("10Y_Treasury_Yield", "2Y_Treasury_Yield", "Fed_Funds_Rate", "CPI", "Industrial_Production", "Unemployment_Rate", "VIX").
2. **Feature Engineering:** 19 features were engineered from a 60-day rolling window, this window was chosen as a balance between stability and responsiveness, limiting noisy data but short enough to capture new regimes. These features were designed to showcase the character of the market’s internal structure and included correlation stats, eigenvalue features, and graph network metrics.
3. **Dimensionality Reduction:** Many of these features were highly correlated, to help with computational efficiency, preventing the model from “double counting” signals, and added noise,” Principal Component Analysis was used to reduce the features into 6 uncorrelated Principle Components




```
Number of components needed for 90% variance: 6
PCA feature matrix shape: (3287, 6)
PCA Explained Variance Ratios: [0.43611283 0.20692335 0.10968091 0.08219468 0.0582279
1 0.03870029]
```

4. **Regime Detection:** A Gaussian Hidden Markov Model was trained on these 6 components to find the 3 hidden states in the market on any given day.



The HMM was able to successfully identify three distinct regimes by analyzing the average value of each state. The model was validated as statistically significant (Chi-Squared p-value < .001).

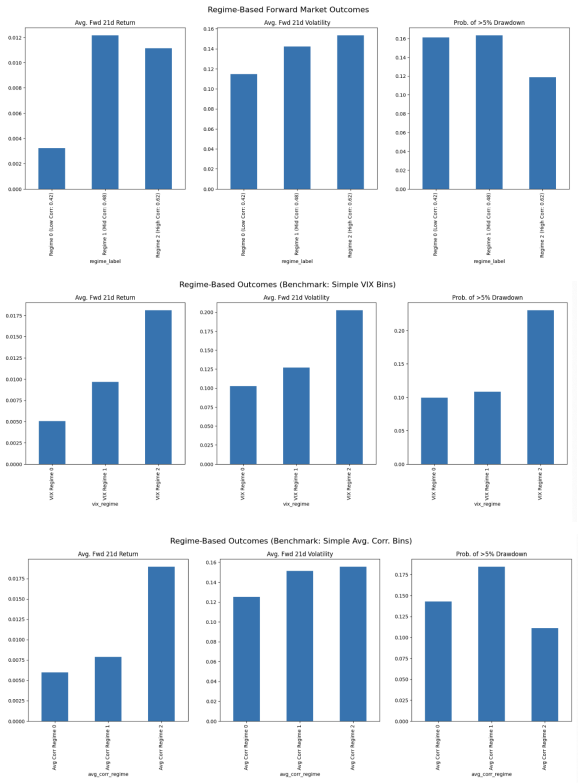
regime	avg_corr	corr_std	top_eig	eig_gap
2	0.422426	0.255649	5.815440	4.347205
1	0.479936	0.198807	6.092250	4.659088
0	0.624721	0.155497	7.412212	6.390496

To test the model’s prediction power, I employed forward looking tests. The model looks at past 60 days of data to characterize the current regime, and then tests its characterization by looking at future 21 day data from SPY, from 2013 to the present. These tests showed:

- **Forward volatility** can be predicted, with Regime 2 (High-Corr) having the highest future volatility (15.4%).
- **Forward Drawdown** cannot be predicted. Regime 2 (High-Corr) had the lowest probability of a > 5% drawdown (11.9%) while the “calm” and “transition” regimes had the highest probability (16.1% and 16.3%)
- **Forward Return** can be predicted. Regime 2 (High-Corr) preceded strong positive returns (1.11%) matching Mid-Corr regimes (1.22%) and exceeding “Low-Corr” regimes (0.32%)

These findings are surprising because it implies high correlation is not necessarily a “risk-off” signal. To understand the benefit of the HMM model, two simpler benchmark models were tested: one based only on VIX and one based only on average correlation.

The VIX benchmark showed the “classic” relationship, a high VIX state correctly predicted low future returns and high drawdown risk. The average correlation produced results nearly identical to the HMM model, with a high ‘avg_corr’ predicting high returns, high volatility, and low drawdown risk.



The model successfully isolated ‘avg_corr’ as the dominant signal. This insight has direct applications for traders and portfolio managers.

1. **As a Strategy Supplement.** This model can help a manager make decisions, when volatility spikes a manager must decide whether to de-risk. This model helps showcase the difference between:
   a. “Bad Volatility”: A high VIX, but a low/mid ‘avg_corr’ (Regime 0 or 1). These states have the highest crash risk, so the best move would be to de-risk.
   b. “Good Volatility”: A high VIX and a high ‘avg_corr’ (Regime 2). The model shows this state is a “Volatile Risk-On” rally with low crash risk. The manager should stay in the trade to capture upside, and can hedge the high volatility with long put options or another short volatility position.
2. **Regime-Based Options Strategy.** Using the model’s prediction for Volatility and Drawdown Risk to select the best options trade for each regime.
   a. HMM = Regime 0 (Low-Corr)
      i. **Findings:** Low Volatility (11.5%) and High Drawdown Risk (16.1%)
      ii. **Strategy:** Buy "Tail-Risk" Hedges.
      iii. **Action:** In this regime, the low volatility means options (like SPY puts) are cheap. Because the drawdown risk is high, the strategy would be to buy OTM SPY put options. This is a cheap way to protect the portfolio from the 16.1% crash risk that the model identifies, even when the market looks calm.
   b. HMM = Regime 1 (Mid-Corr)
      i. **Findings**: High Return (+1.22%) but the Highest Drawdown Risk (16.3%).
      ii. **Strategy**: Bullish with Defined Risk.
      iii. **Action**: You want to capture the upside, but you must protect against the high chance of a drop. The strategy would be to buy SPY call spreads (e.g., buy a call at the money, sell a call further out of the money). This limits your maximum loss if the market turns against you, while still allowing you to profit from the expected positive returns.
   c. HMM = Regime 2 (High-Corr)
      i. **Findings**: High Return (+1.11%), Highest Volatility (15.4%), but Lowest Drawdown Risk (11.9%).
      ii. **Strategy**: Be Bullish and Sell Volatility.
      iii. **Action**: This is the most powerful signal. The model is bullish (high return) and not afraid of a crash (low drawdown risk), but it expects a lot of chop (high vol). The strategy would be to sell ATM SPY put options. This trade collects a high amount of premium (because volatility is high) and has a high probability of success because the model predicts the market is unlikely to crash.
3. **Long/Short Strategy.** Using the ‘avg_corr’ signal, the dominant feature, to decide which kind of long/short strategy to run.
   a. HMM = Regime 0 or 1 (Low/Mid-Corr)
      i. **Finding**: Average correlation is low. This means assets are not moving together, and diversification works well.
      ii. **Strategy**: Run a Sector-Rotation / Long-Short Equity Strategy.
      iii. **Action**: Because sectors are moving independently, you can profit from the "spread" between them. The strategy would be to go Long risk-on sectors (like XLK - Tech, XLY - Consumer Discretionary) and simultaneously go short risk-off sectors (like XLU - Utilities, XLP - Consumer Staples).
   b. HMM = Regime 2 (High-Corr):
      i. **Finding**: Average correlation is very high (0.62). All assets are moving together ("one single trade").
      ii. **Strategy**: Run an Index-Based / Macro Strategy.
      iii. **Action**: In this regime, diversification fails. It's pointless to be long XLK and short XLU, as the model shows they will just move together. The only trade that matters is the macro-level direction. Based on the findings, the strategy would be to go Long SPY (for the positive returns) and potentially go Short TLT (long-term bonds), as this regime also had the highest Term Spread, suggesting a "steepening" yield curve, which is bad for long-duration bonds.