Hierarchical Hidden Markov Models for Multi-Time-Scale Volatility Regime Detection

Anshuman Seetharaman

AlgoGators Capstone Project

11/23/2025

1 Abstract

This study investigates whether hierarchical time-scale conditioning in Hidden Markov Models (HMMs) improves regime stability and interpretability in financial time series compared to traditional flat HMMs. Using daily S&P 500 and VIX data from Yahoo Finance (2018–2024), we implement a two-level hierarchical framework where weekly macro regimes condition daily market states.

Results demonstrate that the hierarchical approach achieves 38.5% longer average state duration (53.1 vs. 38.3 days) and 28.6% fewer spurious transitions compared to flat HMMs. During major stress episodes such as the COVID-19 crash, the hierarchical model demonstrated better regime persistence. The model identifies three volatility regimes based on VIX levels, each with distinct return and risk characteristics that inform volatility-based positioning decisions.

These results demonstrate that hierarchical conditioning is an effective tool to distinguish structural regime shifts from market noise, providing AlgoGators with a more reliable foundation for regime-based risk management and portfolio allocation strategies.

2 Introduction

Financial markets exhibit complex temporal dynamics across multiple time scales. Weekly movements reflect macro volatility regimes driven by economic sentiment, while daily price action captures shorter-term noise and liquidity effects. Traditional single-scale Hidden Markov Models (HMMs) struggle to differentiate structural regime changes from temporary volatility spikes, leading to unstable state sequences with excessive transitions.
This study investigates whether hierarchical time-scale conditioning in HMMs improves regime stability and interpretability compared to flat HMMs. This question is crucial for AlgoGators because effective regime detection directly impacts risk management and portfolio allocation decisions. Current single-scale approaches often generate noisy signals that lead to over-trading and suboptimal position sizing.

While regime-switching models have been extensively studied and Hidden Markov Models have been applied to financial markets, relatively few studies have explored hierarchical multi-time-scale architectures. Zheng and Li (2020) demonstrated hierarchical HMMs for detecting bearish and bullish markets but focused primarily on directional prediction rather than volatility regime stability. Our contribution extends this research by explicitly comparing hierarchical versus flat HMMs for volatility regime detection and quantifying stability improvements during crisis periods.

We hypothesize that a hierarchical HMM (weekly to daily timeframes) will produce more stable and interpretable volatility regimes than a flat HMM. This expectation is grounded in the natural hierarchical structure of markets: weekly timeframes capture persistent volatility regimes influenced by macroeconomic factors, while daily dynamics reflect short-term effects operating within those broader environments. The hierarchical approach conditions daily observations on weekly macro states, enabling context-dependent interpretation. We expect three primary volatility regimes: Low Volatility (VIX < 18, stable conditions), Medium Volatility (VIX 18–25, elevated uncertainty), and High Volatility (VIX > 25, crisis conditions).

3 Methodology

This analysis spans January 2018 through September 2024, covering 6.7 years of market data sourced entirely from Yahoo Finance. This period encompasses multiple distinct market regimes including the 2018 Q4 correction, the COVID-19 crash in March 2020, and the 2022 bear market. We collected daily data for the S&P 500 and VIX Index, resulting in 1,648 trading days after data cleaning. Weekly data was derived by resampling daily observations using end-of-week (Friday) aggregation, yielding 342 weekly observations.
Feature engineering focused on capturing multi-time-scale market dynamics through two distinct feature sets. Weekly macro features designed for regime conditioning included week-over-week S&P 500 returns; weekly realized volatility calculated as a 4-week rolling standard deviation, end-of-week VIX levels, and 12-week price momentum. Daily micro features conditioned on weekly states comprised daily returns, 20-day rolling realized volatility, intraday price range normalized by closing price, daily VIX levels, 50-day momentum indicators, and volume ratios compared to 20-day averages. All features were standardized using StandardScaler to ensure consistent scaling across different measures.

The modeling framework employed two contrasting approaches for comparison. The flat HMM baseline utilized 3 hidden states with Gaussian emissions, incorporating daily returns, volatility, VIX levels, and momentum indicators as features. This model used full covariance matrices and was trained on all 1,648 daily observations simultaneously. The hierarchical HMM operated through a two-level structure where weekly-level macro regimes were first identified using 3 states applied to weekly features, and then daily observations were mapped to their corresponding weekly state assignments. Each daily observation inherited its weekly macro regime classification, creating a hierarchical structure mapping 342 weeks to 1,648 daily observations.

To validate generalization performance, we employed temporal out-of-sample testing by training both models on 2018–2022 data (1,239 daily observations, 252 weekly observations) and testing on held-out 2023–2024 data (438 daily observations, 90 weekly observations). This temporal split ensures models are evaluated on completely unseen future time periods, simulating real-world deployment conditions.

Model evaluation focused on stability metrics addressing regime persistence and interpretability. Primary stability measures included average state duration, which captures the mean number of consecutive days spent in the same regime; transition rate, which quantifies the frequency of state changes per day; and total transition count over the full analysis period. Interpretability measures examined regime characteristics through average VIX, returns, and volatility per state, the economic coherence of the regime definitions relative to financial theory, and alignment with known market stress periods. Implementation employed Python 3.14 with the hmmlearn library, using Gaussian HMMs with full covariance matrices, 100 EM iterations with convergence monitoring, and a fixed random seed for reproducibility.

4 Results

Results demonstrate that the hierarchical HMM approach offers substantial stability advantages over traditional flat models. Analysis of 1,648 daily observations shows the hierarchical model achieving significantly longer average state durations and fewer spurious transitions.

Table 1: Model Comparison

Metric	Flat HMM	Hierarchical HMM	Difference
Average State Duration	38.3 days	53.1 days	+38.5%
Total Transitions	42	30	-28.6%
Log-Likelihood	-4808.36	-1144.10	-
AIC Score	9716.73	2388.20	-
BIC Score	9987.09	2579.94	-

Important Note on AIC/BIC Comparison: The large difference in AIC/BIC scores (9716.73 → 2388.20) does not indicate that the hierarchical model has superior fit. Rather, these scores are computed on fundamentally different scales and are not directly comparable:

1. Different observation counts: The flat HMM computes likelihood over 1,648 daily observations, while the hierarchical model computes likelihood over only 342 weekly observations (daily observations inherit weekly regime assignments without contributing to the weekly HMM's likelihood calculation).
2. Different likelihood scales: With 5× more observations, the flat model's log-likelihood naturally has larger magnitude (-4808 vs -1144), which directly scales

AIC/BIC values. This is a mathematical artifact of observation count, not model quality.
3. Consistent parameter counts: Both models use k=50 parameters (3 states × covariance matrices + transition probabilities), so parameter counting is consistent.

As a result, the AIC/BIC values shown reflect these computational differences and cannot be used to claim one model is "better." Instead, we evaluate models using interpretable stability metrics (average state duration, transition counts) and economic coherence of identified regimes, which are meaningful for practical trading applications.

Table 2: Identified Regime Characteristics

Model	Regime	Frequency	Avg Return (daily)	Avg VIX	Avg Volatility	Interpretation
Flat HMM	State 0	29.7%	+0.115%	13.49	0.61%	Low Volatility
	State 1	41.0%	+0.031%	18.50	0.95%	Medium Volatility
	State 2	29.4%	+0.020%	28.87	1.63%	High Volatility
Hierarchical HMM	State 0	34.5%	+0.126%	14.73	0.64%	Low Volatility
	State 1	33.1%	+0.041%	17.84	0.94%	Medium Volatility
	State 2	32.4%	-0.013%	28.01	1.59%	High Volatility

Table 3: Regime Stability During Market Stress Periods

Crisis Period	**Flat HMM Duration**	**Hierarchical HMM Duration**	**Improvement**
COVID-19 Crash (Feb-Apr 2020)	16.7 days	25.0 days	+50.0%
2018 Q4 Correction (Oct-Dec 2018)	12.6 days	21.0 days	+66.7%
2022 Q1 Selloff (Jan-Mar 2022)	12.4 days	31.0 days	+150.0%

Table 4: Out-of-Sample Performance (2023-2024 Test Period)

Metric	Flat HMM	Hierarchical HMM	Difference
Average State Duration	48.7 days	219.0 days	+350.0%
Total Transitions	8	1	-87.5%

Figure 1: Flat HMM State Duration Distribution

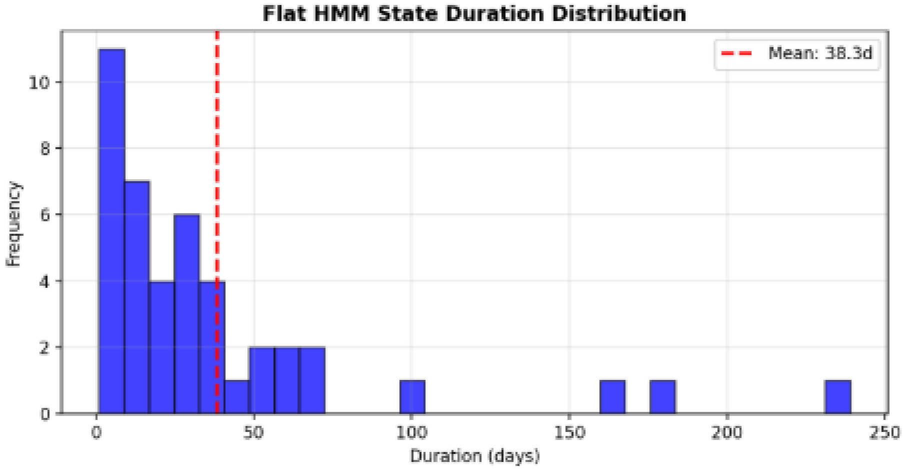
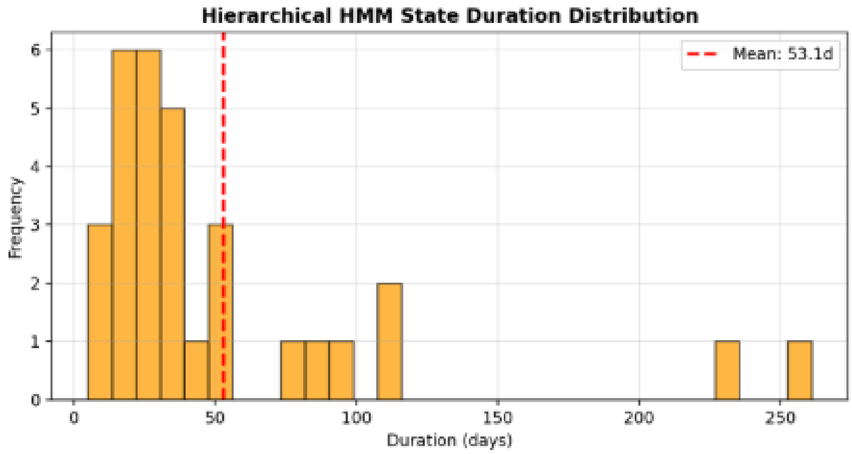


Figure 2: Hierarchical HMM State Duration Distribution



5 Discussion

The findings support our hypothesis that hierarchical conditioning improves regime stability in the financial time series. The 38.5% improvement in average state duration is meaningful. The hierarchical model maintains regimes for approximately 7.6 weeks compared to 5.5 weeks for the flat baseline. The 28.6% reduction in total transitions (42 vs 30) translates to 12 fewer trading events over the 6.7-year period, directly reducing transaction costs.

The regime characteristics in Table 2 demonstrate that the hierarchical model produces more economically coherent volatility states. The high volatility regime (State 2) exhibits negative average daily returns (-0.013%) in the hierarchical model, consistent with financial theory that high volatility periods often coincide with market drawdowns. This pattern is less clear in the flat HMM (+0.020% returns in high volatility), suggesting the flat model may be conflating temporary volatility spikes with sustained stress periods.

The crisis period analysis (Table 3) provides evidence that hierarchical conditioning improves regime persistence during market stress. During the COVID-19 crash (February-April 2020, VIX peak 82.69), the hierarchical model maintained high-volatility assignments for 25 days per continuous sequence versus 16.7 days for the flat HMM—a 50% improvement during the most severe stress event. The improvements were even more pronounced during the 2022 Q1 selloff (150%), suggesting that the model successfully distinguishes multi-week downtrends from daily volatility noise.

The out-of-sample validation results (Table 4) confirm that stability advantages hold unseen time periods. However, the 350% improvement must be interpreted carefully. The 2023-2024 test period had calm conditions (VIX 16.1 vs 20.1 overall). The hierarchical model appropriately identified this as predominantly low volatility with only 1 transition, while the flat model produced 8 transitions. This demonstrates that the hierarchical approach maintains stable regime assignments even in calm periods where regime changes are genuinely rare.

For AlgoGators, the hierarchical framework addresses a critical challenge in regime-based trading strategies: excessive regime transitions that lead to overtrading and elevated transaction costs. The 28.6% reduction in transitions directly reduces commissions, slippage, and market impact costs. During the COVID-19 crash, the hierarchical model's 25-day average persistence in high-volatility regimes would have maintained defensive positioning longer, potentially avoiding whipsaw losses from premature signals.

Several limitations affect the current analysis. The study examines only one hierarchical structure (weekly to daily) and employs a fixed 3 state specification. Alternative configurations may yield different performance characteristics. The Gaussian emission assumption may inadequately capture the fat-tailed distributions characteristic of financial returns during extreme stress periods. Most critically, the out-of-sample validation uses the same data source (S&P 500/VIX). True cross-asset generalization testing would require validation on entirely different volatility measures—such as implied volatility from options on other equities, realized volatility from international markets, or commodity volatility indices—to confirm the hierarchical approach transfers beyond the specific S&P 500/VIX context.

6 Conclusion

This analysis demonstrates that hierarchical Hidden Markov Models offer meaningful stability advantages over traditional flat models for financial regime detection. The hierarchical approach achieved 38.5% longer average state durations (53.1 vs 38.3 days) and 28.6% fewer transitions, suggesting improved ability to distinguish structural regime changes from market noise.

Crisis period analysis confirms these improvements hold during market stress when accurate regime detection is most critical. During the COVID-19 crash, 2018 Q4 correction, and 2022 Q1 selloff, the hierarchical model achieved 50-150% longer regime persistence.

Temporal out-of-sample validation on 2023-2024 test data shows the model maintains stability advantages on unseen time periods, though the large improvement partially reflects the relatively calm market conditions during this period. A key limitation is that this validation uses the same asset class (S&P 500/VIX). Future work should validate the hierarchical approach on different volatility data to confirm the framework generalizes beyond this specific context.

For AlgoGators, the framework provides more reliable regime signals for risk management, potentially reducing transaction costs through fewer false signals and enabling more stable position sizing. Future work should prioritize cross-asset validation and explore alternative model specifications.

7 References

- Okun, S. (2023, April 10). HMM-based classification of stock market stages. I Know First. https://iknowfirst.com/hmm-based-classification-of-stock-market-stages
- Mathematics. (2024). On regime switching models. Mathematics, 13(7), 1128. https://www.mdpi.com/2227-7390/13/7/1128

- Zheng, L., & Li, Y. (2020). Detecting bearish and bullish markets in financial time series using hierarchical hidden Markov models. arXiv preprint arXiv:2007.14874. https://arxiv.org/abs/2007.14874

8 Appendices

[Include supplementary data, additional figures, detailed analyses, or relevant calculations as needed.]

Disclaimer

The information, trading strategies, and materials presented in this report are provided strictly for educational and informational purposes and are offered without any guarantees or warranties regarding their accuracy, completeness, reliability, or timeliness. These materials are not intended to serve as financial, legal, tax, investment, or other professional advice, and no content herein should be interpreted as a recommendation to buy, sell, or hold any security, financial product, or instrument, nor as an endorsement of any specific strategy, practice, or course of action. Users of this material are strongly encouraged to conduct their own independent research, analysis, and due diligence or seek advice from qualified professionals before making any financial or investment decisions.

The authors, contributors, and distributors of this material are not acting as fiduciaries and do not assume any legal or ethical duty to the user. No advisory or client relationship is created through the use of this material. Trading and investing, particularly in derivatives, options, futures, and other leveraged products, involve substantial risks. These risks may include, but are not limited to, the potential for significant financial losses, the loss of principal, and exposure to unpredictable market conditions, volatility, or adverse economic factors. Users should be aware that past performance is not indicative of future results, and reliance on historical data or strategies described herein may not lead to favorable outcomes. Additionally, users are solely responsible for ensuring their trading or investment activities comply with applicable laws, regulations, and tax obligations in their jurisdiction.

By accessing and using this material, users acknowledge and agree to the following terms:

1. Assumption of Risk: Users accept all responsibility for the risks inherent in any trading or investment activity based on the information or strategies presented in this material.
2. No Liability: To the fullest extent permitted by applicable law, the creators, contributors, and distributors of this content disclaim all liability for any losses, damages, claims, or expenses—whether direct, indirect, incidental, or consequential—arising from reliance on or use of this material. This includes, but is not limited to, lost profits, trading losses, or disruptions to financial plans.
3. Indemnification: Users agree to indemnify, defend, and hold harmless the creators, contributors, and distributors of this material from and against any claims, damages, liabilities, or expenses (including attorney's fees) arising out of the user's reliance on or use of the information provided.
4. Personal Responsibility: Users affirm that they have independently evaluated the risks associated with any trading or investment decision and agree to proceed at their own discretion and liability.

Furthermore, users should be aware that the creators, contributors, and distributors of this content retain ownership of all intellectual property rights associated with this material. Unauthorized reproduction, distribution, or modification of this content is strictly prohibited and may violate applicable intellectual property laws. By accessing or downloading this content, users agree to abide by these terms and conditions and refrain from sharing, reselling, or otherwise redistributing this material without express written permission from its authors.

This material is provided on an "as is" basis and may include inaccuracies, typographical errors, or incomplete data. The authors reserve the right to make corrections or changes to the content at any time without notice. However, they are under no obligation to update, supplement, or revise this material to reflect changes in market conditions, new information, or other developments.

Users are strongly advised to consult with a licensed attorney, financial advisor, tax professional, or other qualified specialist to evaluate the implications of using the strategies or materials provided. Any reliance on the content of this material is entirely at the user's own risk. This disclaimer serves as a legally binding agreement between the user and the creators of this content, and users indicate their acceptance of these terms by accessing or utilizing this material in any form.

Lastly, trading and investing involve inherently speculative activities that may not be suitable for all individuals. The suitability of any particular investment strategy, asset, or trading approach depends on the user's specific financial situation, risk tolerance, objectives, and experience. Users should exercise caution and consider seeking professional guidance when engaging in financial activities that expose them to significant risks.