Optimal Portfolio Rebalancing Under Changing Market Conditions

Jessie Dong
Department of Computer Science,
Stanford University
jessiedong@stanford.edu

Antonio Alonso-Stepanov
Department of Computer Science,
Stanford University
ajalonso@stanford.edu

Steve Dou
Department of Computer Science,
Stanford University
stevedou@stanford.edu

Abstract

We address the challenge of portfolio allocation and rebalancing under a market regime that cannot be known with certainty. We implement a Partially Observable Markov Decision Process (POMDP) where the hidden state is the true forward-looking volatility (either calm or volatile), and where observations are the returns for a particular month. The actions are either choosing an aggressive (high stock percentage) or low risk (low stock percentage) allocation. We use the Fama–French 5-factor (2×3) data to first estimate model parameters, and then we compute an optimal policy using value iteration to maximize mean-variance utility (which rewards for higher average returns and penalizes for higher volatility). We compare this POMDP allocation strategy to a simple 60/40 portfolio as well as a volatility threshold portfolio. The POMDP approach yields superior results, providing a higher overall return, a higher Sharpe ratio, and a lower maximum drawdown compared to the alternatives.

I. Introduction

Portfolio rebalancing is a fundamental component of long-term investment management. As asset prices evolve, portfolio weights naturally drift away from their intended allocations. This drift alters the portfolio's risk exposure and can reduce diversification benefits over time. A large body of research shows that unmanaged drift increases volatility, raises downside risk, and causes portfolios to deviate from the investor's strategic objectives [1, 2]. Rebalancing is used to correct these deviations. However, frequent rebalancing incurs avoidable transaction costs, while infrequent rebalancing allows potentially harmful misalignments to accumulate [3]. The central problem is to determine when rebalancing is truly necessary and how to make allocation decisions that balance the tradeoff between maintaining desired exposure and minimizing trading.

Many widely used rebalancing approaches are based on simple heuristics such as fixed schedules or threshold rules [4]. More sophisticated methods treat the problem as an optimization or control task, often assuming that the relevant market information is fully observable. In practice, this assumption is rarely realistic. Financial markets exhibit periods of calm and turbulence, and empirical evidence shows that these periods reflect latent states that cannot be observed directly but influence volatility, returns, and risk premia [5, 6]. Investors must therefore infer underlying conditions from noisy signals. This motivates the need for rebalancing frameworks that explicitly model partial observability and uncertainty about market regimes.

Our research addresses this gap by formulating portfolio rebalancing as a Partially Observable Markov Decision Process. In our framework, the market evolves through latent regimes, such as high-volatility and low-volatility periods. The true regime is not observed. Instead, the agent receives noisy volatility information at each time step and uses Bayes' rule to update a belief distribution over the possible regimes. This belief state provides a sufficient representation of the information relevant for decision-making. At each step, the agent chooses between a high-risk and a low-risk allocation. The expected return of each action depends on the hidden state, which creates a natural tradeoff between exploiting perceived calm conditions and protecting against the possibility of turbulence.

We evaluate this framework using the Fama and French Five Factor (2 by 3) dataset [7], which provides a well-defined structure for modeling return and volatility dynamics. We define the hidden regime, construct an observation process based on factor volatility, and compute the optimal allocation policy using value iteration on belief space. Our results show that the POMDP policy outperforms both a 60/40 static allocation and an active volatility threshold heuristic approach. The POMDP simultaneously achieves the highest total returns, the highest Sharpe ratio, and the lowest maximum drawdown among the tested strategies.

II. Related Works

Prior work on portfolio rebalancing and dynamic allocation has explored both heuristic strategies and optimization-based frameworks. Campbell and Viceira showed that portfolio drift can significantly alter long-run risk exposure, which motivates systematic rebalancing practices [1]. Sharpe emphasized that stable asset allocation is a primary determinant of realized returns for diversified portfolios [2]. These studies establish the importance of maintaining consistent exposure over time.

A large portion of the literature relies on simple rebalancing rules. Fixed-interval and threshold-based strategies remain common due to their ease of implementation and predictable trading patterns [3]. Perold and Sharpe analyzed the effectiveness of such rules under proportional transaction costs and argued that they can work well in environments with limited volatility [4]. More formal approaches model rebalancing as a stochastic control problem. Magill and Constantinides and later Davis and Norman demonstrated that transaction costs naturally lead to regions in which no trading is optimal, since the benefit of correcting small deviations does not outweigh the cost of trading [5, 6]. These results provide important theoretical insights under full observability of market conditions.

A separate line of research investigates structural breaks and regime behavior in financial markets. Hamilton's pioneering work introduced a Markov-switching model for macroeconomic time series and showed that latent regime identification improves predictive accuracy relative to linear models [7]. Ang and Bekaert extended regime-switching ideas to global asset allocation and demonstrated that returns and volatilities vary meaningfully across latent states [8]. Additional studies document that equity markets frequently alternate between calm and turbulent periods that correspond to different risk premia and volatility structures [9]. These findings suggest that market conditions evolve through hidden states that investors must infer indirectly from observable signals.

Recent work has applied reinforcement learning and other machine learning techniques to dynamic allocation. Several studies treat portfolio management as a sequential decision problem, although most assume full observability of the environment or rely directly on price-based features without modeling latent state structure [10, 11]. Partial observability has received relatively limited attention in the context of rebalancing, despite strong empirical support for regime-driven financial dynamics.

Our work contributes to this literature by combining belief-based inference with dynamic allocation decisions. The POMDP framework provides a natural way to represent hidden market regimes, incorporate noisy volatility information, and compute optimal policies when the underlying state cannot be observed directly. This approach connects regime-switching time series models with decision-making under uncertainty and offers a new perspective on portfolio rebalancing in environments where market conditions must be inferred rather than observed.

III. Data

We used the Fama-French 5 Research Factors (2×3) dataset from the Kenneth French Data library [11], which is one of the most widely used resources in asset pricing research. This dataset provides monthly return data for factor-based portfolios constructed from all NYSE, AMEX, and NASDAQ stocks with all the necessary data we need.

The dataset is formatted as monthly observations of the key factors we used in asset pricing models. The dataset includes the Mkt–RF, SMB, HML, RMW, and CMA factors, along with the one-month Treasury bill rate (RF):

- Mkt-RF: The excess return of the market portfolio over the risk-free rate
- SMB: The size premium, which captures returns of small-cap versus large-cap stocks
- HML: The value premium
- RMW: The profitability premium, based on operating profitability
- CMA: The investment premium, based on asset growth
- RF: The risk-free rate (one-month Treasury bill rate)

Preprocessing We restrict our analysis to the period from January 2010 onward, giving us about 15 years worth of data for our model estimation and simulation. This time period captures multiple market events, including the post-2008 financial crisis recovery, periods of low volatility, the COVID-19 market conditions in 2020, and subsequent market conditions. To obtain a single observable return series and estimate forward-looking market volatility, we use the market's monthly total return and compute rolling volatility.
Months with volatility below the median were labeled as "calm" regimes, while those above the median were labeled as "volatile" regimes. These serve as what we used for the ground truth for estimating our POMDP transition probabilities but they were treated as "unobservable" during our policy execution. Thus, our agent must infer the current regime from noisy observations. This thresholding gives two regimes with nearly even representation (92 calm months vs. 97 volatile months), reducing class imbalance and allowing us to do stable transition estimation. As shown in Figure 1, the estimated transition probabilities indicate that both calm and volatile regimes are highly persistent, while still allowing for meaningful transition rates between states. The nearly balanced regime distribution further supports that we can do reliable estimation of regime dynamics.

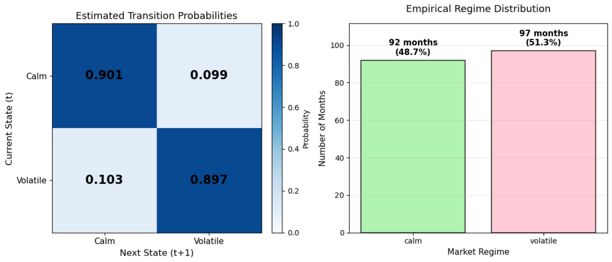


Figure 1. Regime transition probabilities (left) and empirical regime distribution (right).

Figure 2 illustrates how these inferred regimes correspond to observed market behavior. Volatile regimes align with clusters of large-magnitude returns and elevated rolling volatility during well-known periods such as 2011, 2018, 2020, and 2022, whereas calm regimes correspond to extended periods of low volatility, particularly from 2013 to 2017. While volatility spikes are observable once future data is observed, the underlying regime is not directly observed in real time, motivating our treatment of the regime as a hidden state in the POMDP.

Overall, this dataset and our chosen subset are particularly well-suited for our problem because we have reliable, professionally-created return series from the CRSP (Center for Research in Security Prices) and Ibbotson.

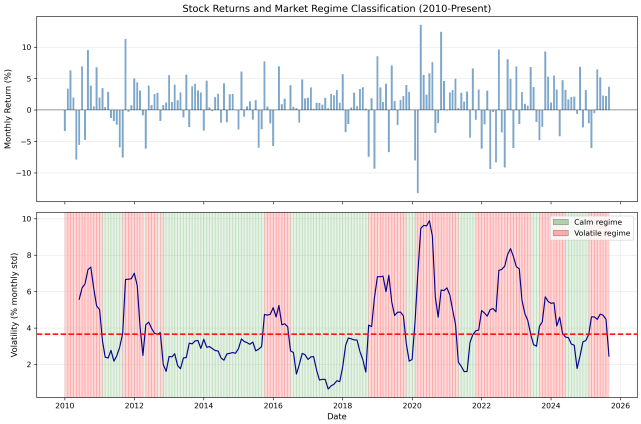


Figure 2. Monthly returns (top) and rolling volatility with inferred calm and volatile regimes (bottom).

IV. Methodology

We model this problem as a Partially Observable Markov Decision Process (POMDP) where the true risk regime is hidden at any given time. To define a POMDP we must specify (S, A, T, R, O, γ). S is the hidden market regime; in particular there are two states: calm (low future variance) and volatile (high future variance) over the next four months. The set of actions are to allocate toward a risky portfolio (90% stocks, 10% cash) or a safe portfolio (35% stocks, 65% cash). The observation is the magnitude of the current monthly return; we discretize this into a crash (a large negative return of less than -0.5%) or a normal return ($> -0.5\%$). The transitions are the likelihoods of moving between calm and volatile market regimes; this is estimated from the historical data. The reward is the risk-adjusted return for that month. More specifically, we use the following formula which has a risk penalty defined by λ, which we set to 10.

$$R(s,a) \;=\; \mathbb{E}\left[r_{s,a}\right] \;-\; \frac{\lambda}{2}\,\mathrm{Var}\left(r_{s,a}\right)$$

So, higher mean return increases the reward, and higher variance decreases the reward. We used a gamma value of 0.99 so that we prioritize long-term gains over short-term returns.

To get the optimal policy, we used model-based value iteration. First, we estimated the parameters T and O from the historical data. We do this by labeling months as volatile or calm based on forward-looking volatility over the following four months and then counting how often one regime transitions to another and normalizing (to get T). We also count how often each regime produces a crash vs a normal return and normalize (to get O). Then, we use the Bellman Optimality Equation and perform updates until our value function converges and we can get a stable optimal policy.

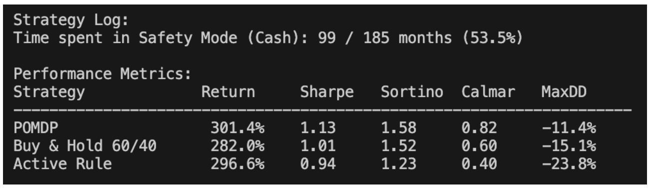
$$V_{k+1}(b) = \max_{a \in A} \left[R(b,a) + \gamma \sum_{o \in O} P(o \mid b, a) V_k(b') \right]$$

Outperforming the benchmarks required a lot of experimentation and tuning of parameters. For instance, the threshold we used for determining if the returns we observed constituted a crash or not required tuning (in other words for discretizing the observations) needed to strike a balance between being too strict (-1.0%), in which case we'd not raise enough alarms, and too loose (-0.2%) in which case we would raise too many alarms about normal volatility. Another aspect that required tuning was the allocations used in the two possible actions. Initially, the risky portfolio was 100% stocks and the safe portfolio was 100% cash. However, to beat both alternative strategies along all of the performance metrics, we needed to tune this; what ended up working was aggressive being 90% stocks and safe being 35% stocks.

The first baseline strategy is a simple buy & hold 60/40 split, which simply involves buying 60% stocks 40% cash at the start of the period and holding onto that for the entire duration. The other strategy is an active rule, which after seeing a "crash" observation (less than -0.5% returns) switches to a safe portfolio and after seeing a "normal" (non-crash) observation (greater than -0.5% returns) switches to an aggressive portfolio.

V. Results

The POMDP strategy delivered stronger risk-adjusted performance than both the 60/40 allocation and the active volatility rule. Across the 2010 to 2025 period, the POMDP achieved the highest total return, Sharpe ratio, Sortino ratio, and Calmar ratio while also producing the smallest maximum drawdown. It grew the portfolio to roughly four times its initial value and maintained a more stable compounding path. The 60/40 portfolio had smoother early growth but lagged in total return, and the active rule showed higher peaks in strong markets but suffered deeper drawdowns.



```
Strategy Log:
Time spent in Safety Mode (Cash): 99 / 185 months (53.5%)

Performance Metrics:
Strategy              Return     Sharpe   Sortino  Calmar   MaxDD
-----------------------------------------------------------------------
POMDP                 301.4%     1.13     1.58     0.82     -11.4%
Buy & Hold 60/40      282.0%     1.01     1.52     0.60     -15.1%
Active Rule           296.6%     0.94     1.23     0.40     -23.8%
```

Figure 3. Performance metrics and cumulative wealth for the POMDP strategy, 60/40 buy-and-hold, and the active rule.

The cumulative wealth paths show that the POMDP tracks market gains during extended calm periods yet avoids the largest declines during volatility spikes. This behavior arises from how the POMDP adjusts exposure based on its inferred hidden regime rather than reacting only to observed returns.

Figure 5 shows how the POMDP's wealth path aligns with its belief about being in a calm regime. The belief varies between about 0.3 and 0.7 and reacts quickly after negative return observations. The agent increases its probability of turbulence before major drawdowns, which allows it to reduce risk in advance.

The drawdown behavior emphasizes the benefit of belief-driven allocation shifts. The POMDP avoids the deepest losses that the 60/40 portfolio experiences during major turbulence episodes in 2011, 2018, 2020, and 2022. Its maximum drawdown of -11.4% is smaller

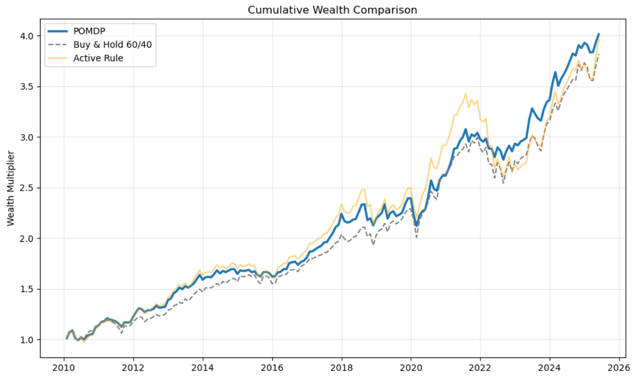


Figure 4. Wealth over time with different strategies

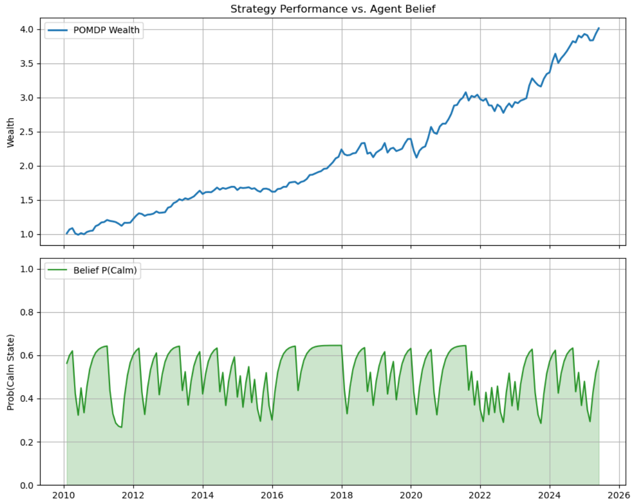


Figure 5. Top: POMDP cumulative wealth. Bottom: posterior belief that the market is in the calm regime.

than the 60/40 portfolio's −15.1% and far below the active rule's −23.8%.

Taken together, these results show that modeling rebalancing as a partially observable control problem produces meaningful improvements over static allocation and simple heuristics. The POMDP adapts to latent regime shifts, balances growth and protection, and generates stronger long-term outcomes by using inference rather than reacting to observed returns.

VI. Conclusion

Our research indicates that modeling the problem of portfolio rebalancing under uncertain market conditions as a POMDP improves risk-adjusted returns, improves total returns, and limits max drawdowns compared to baseline heuristic approaches, including a 60/40 portfolio and a simple volatility threshold strategy. We define the hidden state as the uncertain forward-looking market volatility and treat monthly returns as noisy observations that give some insight into the state. The POMDP managed to balance growth during calm periods and minimizing downside during volatile periods.

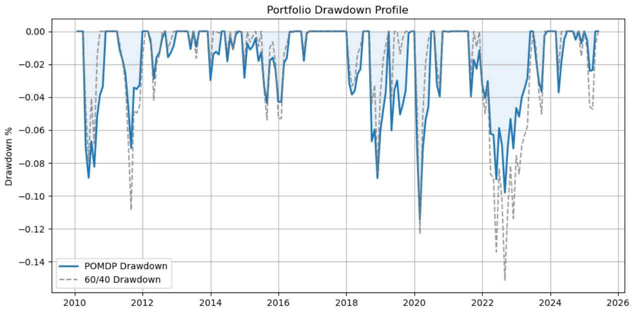


Figure 6. Drawdown profile for the POMDP strategy compared to the 60/40 portfolio.

VII. Future Work

While our POMDP framework demonstrates promising results for portfolio rebalancing under partial observability, there are several possible extensions that could enhance both its applicability and theoretical basis.

Richer State Space and Multi-Regime Models: Our current implementation uses a binary regime classification (calm versus volatile). Future work could explore more granular regime structures, such as three or more states that capture intermediate market conditions like moderate volatility, momentum-driven rallies, or crisis periods. This would allow the model to differentiate between subtly and harder to differentiate market environments and potentially respond with more specific allocation strategies. Additionally, incorporating regime-switching models with time-varying transition probabilities could better capture the empirical observation that market dynamics themselves evolve over longer horizons.

Enhanced Observation Models: Similarly, we currently discretize observations into "crash" and "normal" returns based on a simple threshold rule. Future research could develop more sophisticated observation functions that incorporate multiple signals simultaneously, such as realized volatility, trading volume, credit spreads, or option-implied volatility. This way, we can gather richer information about the hidden state. Other machine learning techniques could be employed to learn optimal observation mappings from high-dimensional market data, potentially from other datasets, which might improve the agent's ability to infer regimes accurately.

Alternative Risk Preferences: We currently optimize mean-variance utility with a fixed risk aversion parameter. Future work should explore other preference structures, such as prospect theory-based utilities that capture asymmetric responses to gains and losses, or conditional value-at-risk (CVaR) objectives that have been mentioned in previous research but don't explicitly target tail-risk management. Incorporating time-varying risk aversion that responds to recent portfolio performance or macroeconomic conditions could also make the strategy more realistic if applied to real-time market data.

VII. Contributions

All team members collaborated closely on the core components of the project, including the design of the POMDP formulation, model estimation, the value-iteration solver, and the overall data analysis. We jointly developed the modeling approach and met weekly to review progress, debug implementation issues, and evaluate experimental outputs. Each member contributed several references to support the literature review, and we worked together to maintain consistency in formatting and citations on Overleaf. In terms of individual responsibilities, Steve led the development of the data section and contributed to the future work and conclusion. Jessie wrote the introduction and related works sections and also contributed to the conclusion. Antonio implemented the methodology and produced the results analysis.

After the checkpoint, Steve and Antonio spent an extra 30 hours redesigning the hidden-state representation from a trailing volatility score to a forward-looking volatility regime model. We updated the POMDP formulation and implementation and did additional testing to ensure that we outperformed our baselines with this more realistic formulation (the 60/40 and volatility threshold portfolios). This required extra data processing as well as debugging when the model initially did not perform up to our expectations.

VIII. References

[1] Campbell, John Y. and Viceira, Luis M. Strategic Asset Allocation. Oxford University Press, 2002. https://global.oup.com/academic/product/strategic-asset-allocation-9780198296942

[2] Sharpe, William F. "Asset Allocation: Management Style and Performance Measurement." Journal of Portfolio Management, 1992. https://jpm.pm-research.com/content/18/2/7

[3] Perold, André and Sharpe, William. "Dynamic Strategies for Asset Allocation." Financial Analysts Journal, 1988. https://www.jstor.org/stable/4479226

[4] Pliska, Stanley R. Introduction to Mathematical Finance. Blackwell Publishers, 1997. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470317083

[5] Ang, Andrew and Bekaert, Geert. "International Asset Allocation with Regime Shifts." Review of Financial Studies, 2002. https://doi.org/10.1093/rfs/15.4.1137

[6] Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." Econometrica, 1989. https://users.ssc.wisc.edu/ behansen/718/Hamilton1989.pdf

[7] Fama, Eugene F. and French, Kenneth R. "A Five-Factor Asset Pricing Model." Journal of Financial Economics, 2015. https://doi.org/10.1016/j.jfineco.2014.10.010

[8] Maheu, John M. and Gordon, Stephen. "Learning, Forecasting and Structural Breaks." Journal of Applied Econometrics, 2008. https://doi.org/10.1002/jae.1002

[9] Moody, John and Saffell, Matthew. "Learning to Trade via Direct Reinforcement." IEEE Transactions on Neural Networks, 2001. https://doi.org/10.1109/72.935097

[10] Jiang, Zhengyao, Xu, Dixing, and Liang, Jinjun. "A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem." ArXiv preprint, 2017. https://arxiv.org/abs/1706.10059

[11] French, K. R. (2025). Data library [Data set]. Tuck School of Business at Dartmouth. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/da