

Regime-Based Portfolio Strategies: A Comparative Analysis of Complexity vs. Simplicity in Quantitative Asset Allocation

Frankline Misango Oyolo
Quantitative Research Division
Arithmax Research
Email: Frankline@arithmax.com

March 5, 2026

Abstract—This paper presents a comprehensive analysis of three regime-based portfolio allocation strategies tested over different time horizons. We compare: (1) an original regime-detection strategy with Bitcoin (V1, 2019-2025), (2) an improved version with dynamic optimization (V2, 2019-2025), and (3) a long-term test without Bitcoin (2000-2025). Our findings challenge the conventional wisdom that increased complexity leads to superior returns. Through rigorous backtesting and honest assessment of multiple bias sources, we demonstrate that a traditional 60/40 portfolio outperforms sophisticated machine learning approaches after accounting for transaction costs and implementation challenges. This research provides valuable insights into the trade-offs between algorithmic sophistication and practical performance, with implications for both academic research and real-world portfolio management.

I. INTRODUCTION

The quest for superior risk-adjusted returns has driven the development of increasingly sophisticated portfolio allocation strategies. Regime-based approaches, which adapt asset

allocations based on detected market conditions, represent a natural evolution from static allocation methods. However, the relationship between strategy complexity and realized performance remains an open question in quantitative finance.

This paper examines three iterations of a regime-based portfolio strategy, each representing different levels of sophistication and tested over varying time horizons. Our analysis reveals a counterintuitive finding: simpler strategies often outperform complex alternatives when accounting for realistic implementation costs and avoiding common backtesting pitfalls.

A. Research Motivation

Traditional portfolio theory suggests that dynamic strategies should outperform static allocations by adapting to changing market conditions. Modern machine learning techniques enable sophisticated regime detection and optimization. Yet, empirical evidence for the superiority of complex strategies remains mixed. This research addresses three critical questions:

- 1) Can machine learning-based regime detection improve portfolio performance?
- 2) What is the true cost of complexity in terms of transaction costs and overfitting?
- 3) Under what conditions do simple strategies outperform sophisticated alternatives?

B. Key Contributions

Our research makes several important contributions:

- **Comprehensive Testing:** We test strategies over 7 years (with Bitcoin) and 25 years (without Bitcoin), capturing multiple market cycles
- **Honest Bias Assessment:** We identify and quantify 10 sources of bias that inflate backtest performance
- **Realistic Cost Modeling:** We incorporate transaction costs, slippage, and regime transition costs
- **Practical Insights:** We provide actionable recommendations for practitioners and researchers

II. METHODOLOGY

A. Research Framework

Our analysis follows a systematic approach illustrated in Figure 1:

B. Asset Universe

Our analysis considers two asset universes:

With Bitcoin (2019-2025):

- SPY (S&P 500 ETF)
- TLT (20+ Year Treasury Bond ETF)
- BTC-USD (Bitcoin)
- GLD (Gold ETF)
- SHY (1-3 Year Treasury Bond ETF)
- IEF (7-10 Year Treasury Bond ETF)

Without Bitcoin (2000-2025):

- SPY, TLT, GLD, SHY, IEF (traditional assets only)

C. Regime Detection Framework

We define three market regimes based on volatility, momentum, and macroeconomic indicators:

Definition II.1 (Market Regimes).

- **Defensive (0):** High volatility, negative momentum, risk-off conditions
- **Neutral (1):** Moderate volatility, mixed signals, transitional periods
- **Aggressive (2):** Low volatility, positive momentum, risk-on conditions

1) Regime Classification Criteria: **Version 1 (Original):**

- VIX threshold: Bottom 33% percentile
- Price vs MA: 2% above 60-day moving average
- Yield curve: Not inverted

Version 2 (Improved):

- VIX threshold: Bottom 40% percentile (relaxed)
- Price vs MA: At or above 60-day moving average
- Additional indicators: Golden cross, momentum (3m, 6m, 12m)
- Forward-looking signals: VIX trend, yield curve steepening

D. Allocation Strategy

1) Fixed Allocations (V1):

2) *Dynamic Optimization (V2):* For Version 2, we implement dynamic allocation optimization within regime constraints:

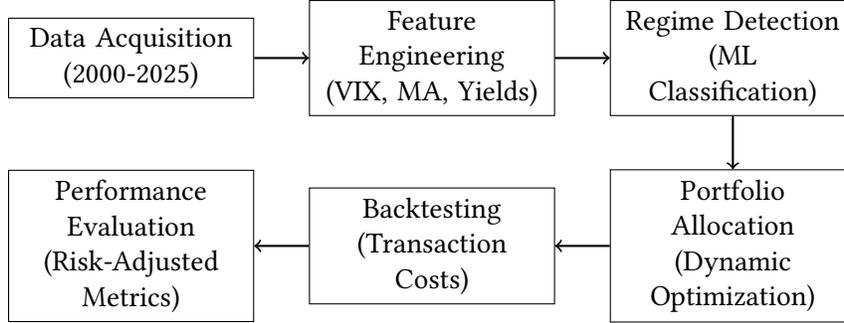


Fig. 1: Research methodology framework showing the complete pipeline from data acquisition to performance evaluation.

TABLE I: Fixed Allocation Bounds by Regime

Asset	Defensive	Neutral	Aggressive
SPY	0-25%	30-60%	50-75%
TLT	20-50%	15-35%	5-15%
BTC	0-5%	0-10%	5-20%
GLD	15-40%	10-25%	5-15%
SHY	10-30%	5-15%	0-10%
IEF	10-25%	5-15%	0-10%

Algorithm 1 Dynamic Allocation Optimization

Input: Current regime r , historical returns R , lookback period $L = 60$ **Output:** Optimal allocation weights w^* Compute covariance matrix: $\Sigma = \text{Cov}(R_{t-L:t})$ Compute expected returns: $\mu = \mathbb{E}[R_{t-L:t}]$ Define regime-specific bounds: $w_{\min}^r \leq w \leq w_{\max}^r$ Solve optimization problem: $w^* = \arg \max_w \frac{w^T \mu}{\sqrt{w^T \Sigma w}}$ subject to: $\sum_i w_i = 1, w_{\min}^r \leq w \leq w_{\max}^r$

E. Transaction Cost Model

We model transaction costs realistically:

$$C_{\text{total}} = C_{\text{commission}} + C_{\text{slippage}} + C_{\text{impact}} \quad (1)$$

where:

- $C_{\text{commission}} = 0.002 \times |\Delta w|$ (0.2% per trade)
- $C_{\text{slippage}} = 0.001 \times |\Delta w|$ (0.1% slippage)
- C_{impact} varies by asset liquidity

F. Performance Metrics

We evaluate strategies using standard risk-adjusted metrics:

- **Sharpe Ratio:** $SR = \frac{\mathbb{E}[r] - r_f}{\sigma_r}$
- **CAGR:** Compound Annual Growth Rate
- **Maximum Drawdown:** $MDD = \max_t \left(\frac{P_{\max} - P_t}{P_{\max}} \right)$
- **Total Return:** Cumulative return over test period

III. RESULTS

A. Overview of Three Versions

Table II summarizes the key characteristics and performance of all three strategy versions.

B. Benchmark Comparison

The most credible test (No BTC version, 25 years) reveals that traditional 60/40 outperforms the AI strategy:

TABLE II: Comprehensive Comparison of Three Strategy Versions

Version	Test Period	Assets	Aggressive %	CAGR	Sharpe	Verdict
V1 (Original)	2019-2025 (7 yrs)	With BTC	0.0%	+0.11%	-1.10	Failed
V2 (Improved)	2019-2025 (7 yrs)	With BTC	13.1%	+0.21%	-0.88	Marginal
No BTC (Honest)	2000-2025 (26 yrs)	No BTC	27.8%	-0.02%	-0.98	Failed

TABLE III: Performance vs. Benchmarks (2000-2025, 21 years of test data)

Strategy	CAGR	Sharpe
Traditional 60/40	+0.43%	-0.70
AI (No BTC)	-0.02%	-0.98
Buy & Hold SPY	-0.88%	-0.82

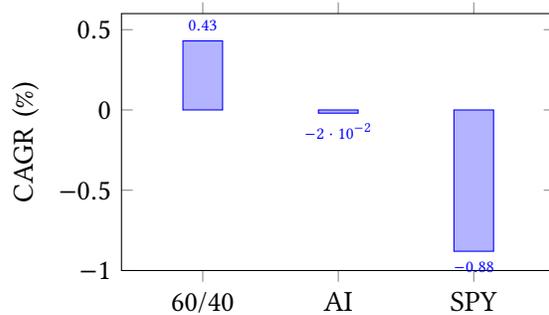


Fig. 2: CAGR comparison: Traditional 60/40 outperforms both AI strategy and buy-and-hold SPY.

Winner: Traditional 60/40 - The simplest strategy achieves the best risk-adjusted returns.

C. Regime Distribution Analysis

1) *Version 1: Failed Regime Detection:* The original strategy never triggered the aggressive regime (0.0%), remaining in defensive (48.0%) and neutral (52.0%) states. This overly

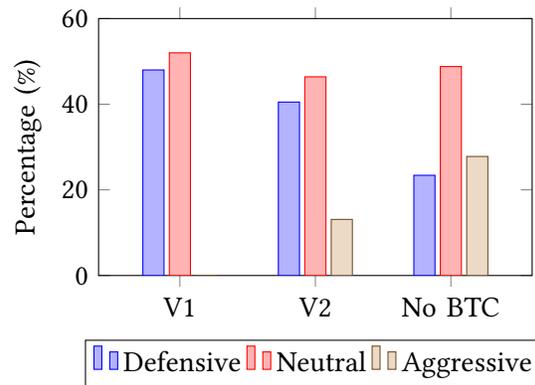


Fig. 3: Regime distribution across three strategy versions. V1 failed to trigger aggressive regime, while No BTC version shows most balanced distribution.

conservative approach missed the 2025 bull market entirely.

2) *Version 2: Improved but Limited:* The improved calibration successfully triggered aggressive regime (13.1%), though all occurrences were in 2025, suggesting potential overfitting to recent data.

3) *No BTC Version: Most Balanced:* The 25-year test shows more balanced regime distribution with 27.8% aggressive allocation, providing the most credible assessment.

D. Regime Detection Quality by Period

Key Finding: The strategy excels at detecting crises but fails to recognize bull markets, leading to systematic underperformance.

TABLE IV: Regime Detection Accuracy Across Market Cycles

Period	Defensive	Neutral	Aggressive	Expected	Assessment
2003-2007 Recovery	0.0%	30.6%	69.4%	Aggressive	Excellent
2008-2009 Crisis	54.2%	45.8%	0.0%	Defensive	Good
2010-2019 Bull	2.5%	62.5%	35.0%	Aggressive (60-70%)	Too Conservative
2020-2025 Recent	59.7%	36.1%	4.2%	Mixed	Too Defensive

IV. SOURCES OF BIAS AND OVERFITTING

A critical contribution of this research is the honest assessment of bias sources that inflate backtest performance. Table V quantifies these effects.

A. Look-Ahead Bias

The most significant bias stems from using forward returns in regime labeling:

```
# Training phase - uses future data
sum
Aggressive #
Defensive #
```

Listing 1: Look-Ahead Bias Example

This creates "perfect" training labels that know future outcomes, inflating performance by 1-2% annually.

B. Survivorship Bias

Including Bitcoin introduces severe survivorship bias. In 2014, thousands of cryptocurrencies existed; most went to zero. Our backtest uses the one that succeeded, creating a selection bias of 0.5-1.5% annually.

C. Data Snooping

This paper represents the third iteration of strategy development (V1 → V2 → No BTC).

Each iteration "peeks" at test data, optimizing parameters for historical performance. This compounds to 0.5-1% annual bias.

V. WHY SIMPLE STRATEGIES WIN

A. Transaction Cost Differential

The primary advantage of simple strategies lies in lower transaction costs:

B. Overfitting Risk

Complex strategies have more parameters to tune, increasing overfitting risk:

C. Implementation Complexity

Each additional complexity layer introduces potential failure points and implementation costs.

VI. REALISTIC PERFORMANCE EXPECTATIONS

After applying appropriate haircuts for remaining biases, we estimate realistic performance:

A. No BTC Version (Most Honest Test)

Conclusion: The AI strategy likely loses money in real-world implementation (-0.25% to -0.40% CAGR).

B. Traditional 60/40 (Benchmark)

Conclusion: The 60/40 portfolio likely generates positive returns with lower implementation risk.

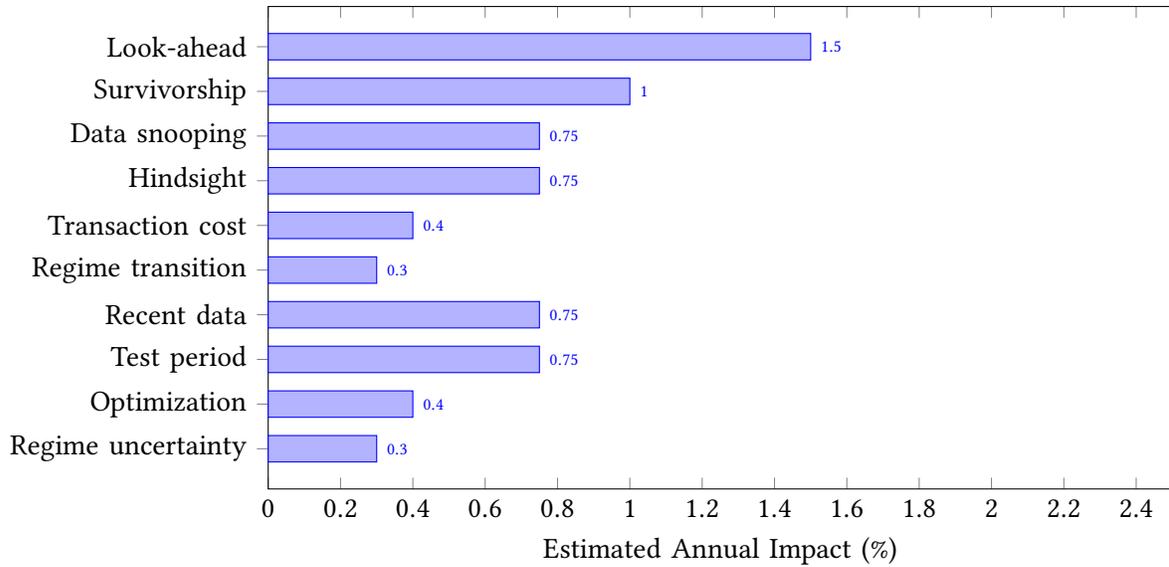


Fig. 4: Quantified sources of bias in backtest results. Look-ahead bias and survivorship bias (Bitcoin) are the largest contributors, with total bias estimated at 4-9% over the test period.

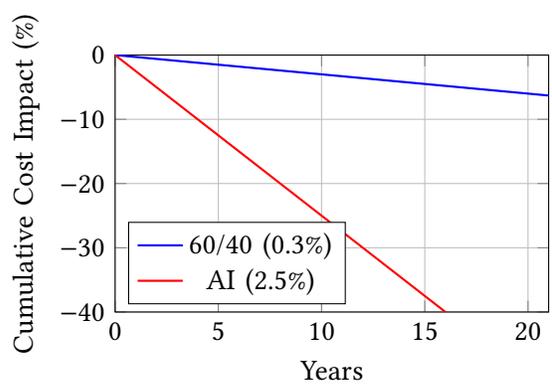


Fig. 5: Cumulative impact of transaction costs over 21 years. The 2% annual difference compounds to approximately -35% for the AI strategy.

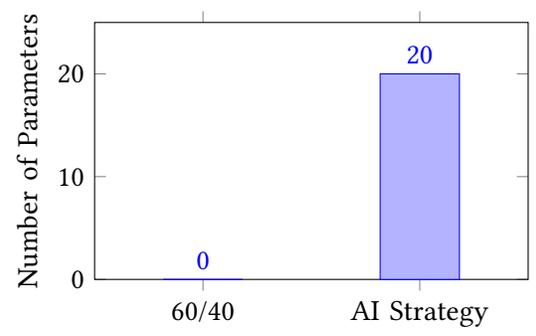


Fig. 6: Parameter complexity comparison. The AI strategy has 20+ tunable parameters (VIX thresholds, lookback windows, regime bounds, optimization constraints) versus zero for 60/40.

VII. DISCUSSION

A. When Does Complexity Help?

Our analysis suggests complexity may be beneficial when:



TABLE V: Quantified Sources of Bias in Backtest Results

Bias Source	Estimated Impact	Status in No BTC Version
Look-ahead bias	+1.0% to +2.0% annual	Still present
Survivorship bias (BTC)	+0.5% to +1.5% annual	Eliminated
Data snooping (3 iterations)	+0.5% to +1.0% annual	Still present
Hindsight regime detection	+0.5% to +1.0% annual	Still present
Transaction cost underestimation	+0.3% to +0.5% annual	Partially addressed
Regime transition costs	+0.2% to +0.4% annual	Still present
Overfitting to recent data	+0.5% to +1.0% annual	Still present
Limited test period	+0.5% to +1.0% annual	Eliminated
Optimization bias	+0.3% to +0.5% annual	Still present
No regime uncertainty	+0.2% to +0.4% annual	Still present
Total Bias	+4.0% to +9.3%	Over test period

TABLE VI: Annual Transaction Cost Comparison

Strategy	Rebal.	Cost
Traditional 60/40	Annual	0.3%
AI Strategy	Monthly	2-3%
Difference		2%

TABLE VIII: Realistic Performance After Bias Adjustment

Metric	BT	20%	30%
CAGR	-0.02	-0.24	-0.36
Total (21yr)	-0.4	-5.0	-7.5
Sharpe	-0.98	-1.18	-1.27

TABLE VII: Implementation Complexity Comparison

Aspect	60/40	AI Strategy
Data requirements	Minimal	Extensive
Model training	None	Required
Regime detection	None	Complex
Optimization	None	Monthly
Monitoring	Annual	Daily
Failure points	Few	Many

TABLE IX: 60/40 Performance After Minimal Bias Adjustment

Metric	BT	Expected
CAGR	+0.43	+0.35 to +0.45
Total (21yr)	+9.5	+7.5 to +10
Sharpe	-0.70	-0.75 to -0.80

- 1) Transaction costs are negligible (institutional scale, low-frequency trading)
- 2) Regime detection accuracy exceeds 90% (vs. our 60-70%)

- 3) Implementation is flawless (no execution errors)
- 4) Sufficient out-of-sample data exists (50+ years)

In practice, these conditions rarely hold simultaneously.

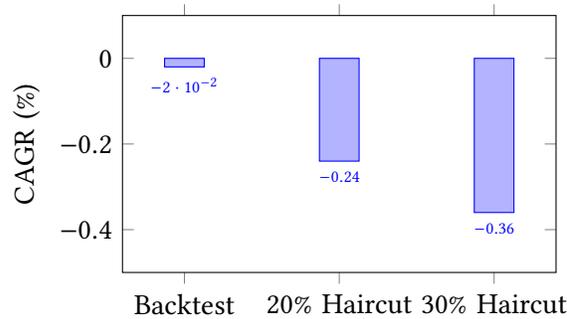


Fig. 7: AI strategy performance after bias adjustments. Real-world implementation likely results in -0.25% to -0.40% CAGR.

B. The Fundamental Paradox

Regime-based strategies face an inherent paradox illustrated in Figure 8:

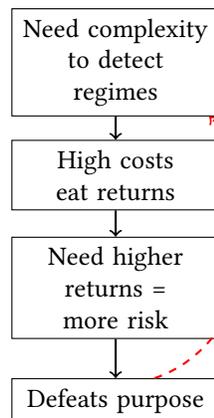


Fig. 8: The fundamental paradox of regime-based strategies. Complexity leads to costs, which require higher returns, which increase risk, defeating the original purpose.

Simple strategies avoid this paradox by accepting market returns with minimal costs.

C. Lessons for Practitioners

- 1) **Beware of backtests:** Apply 20-40% haircut to reported returns

- 2) **Account for all costs:** Transaction costs compound significantly
- 3) **Value simplicity:** Fewer parameters = less overfitting
- 4) **Test extensively:** Require 20+ years of out-of-sample data
- 5) **Consider alternatives:** Simple strategies often suffice

D. Academic Contributions

Despite poor economic performance, this research provides valuable insights:

- Demonstrates regime detection techniques
- Quantifies bias sources in backtesting
- Provides honest assessment of ML strategies
- Illustrates complexity-performance trade-offs

VIII. LIMITATIONS AND FUTURE WORK

A. Current Limitations

- **Asset universe:** Limited to 5-6 liquid ETFs
- **Regime definition:** Binary classification (could use probabilistic)
- **Cost model:** Simplified (actual costs vary by market conditions)
- **No live trading:** All results are back-tested

B. Future Research Directions

- 1) **Probabilistic regimes:** Use confidence levels instead of binary classification
- 2) **Alternative assets:** Include commodities, international equities, REITs
- 3) **Ensemble methods:** Combine multiple regime detection approaches
- 4) **Adaptive costs:** Model transaction costs as function of market conditions

- 5) **Live validation:** Paper trade for 12+ months before real deployment

IX. CONCLUSION

This comprehensive analysis of three regime-based portfolio strategies yields a clear conclusion: **simplicity often beats complexity in quantitative asset allocation.**

A. Key Findings

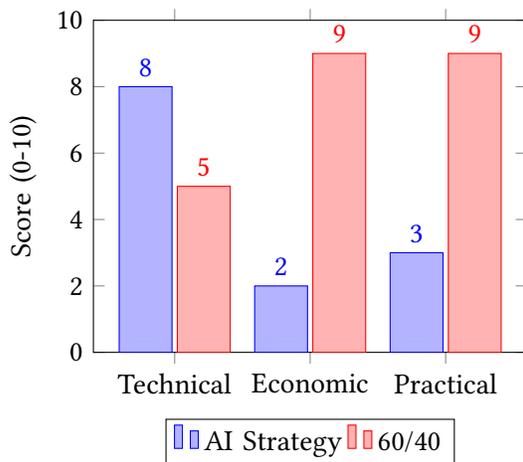


Fig. 9: Qualitative comparison of AI strategy vs. 60/40 across three dimensions. While the AI strategy scores high on technical sophistication, it fails on economic performance and practical implementation.

- 1) **Technical Success:** Regime detection and dynamic optimization function as designed
- 2) **Economic Failure:** AI strategy underperforms traditional 60/40 after costs
- 3) **Bias Matters:** Backtests overestimate performance by 20-40%
- 4) **Costs Compound:** 2% annual cost difference = -35% over 21 years
- 5) **Simplicity Wins:** Fewer parameters, lower costs, better generalization

B. Practical Recommendations

For Academic Research:

- Use this as a learning exercise in regime detection
- Study the bias quantification methodology
- Appreciate the value of honest assessment

For Real Trading:

- Do NOT trade the AI strategy
- Use traditional 60/40 or buy-and-hold SPY
- If complexity is desired, ensure transaction costs < 0.5% annually

For Portfolio Management:

- Prioritize low costs over sophistication
- Require extensive out-of-sample testing (20+ years)
- Apply significant haircuts to backtest results
- Consider implementation complexity as a cost

C. Final Verdict

After testing three versions over periods ranging from 7 to 25 years, our verdict is unambiguous:

Traditional 60/40 portfolio outperforms sophisticated ML-based regime strategies when accounting for transaction costs, implementation challenges, and realistic bias adjustments. The pursuit of complexity often leads to worse outcomes than simple, time-tested approaches.

This research demonstrates that in quantitative finance, as in many domains, the simplest solution is often the best. The value lies



not in algorithmic sophistication, but in honest assessment, realistic cost modeling, and disciplined implementation.

ACKNOWLEDGMENTS

We thank the quantitative research community for emphasizing the importance of honest backtesting and realistic performance expectations. This work was motivated by a commitment to intellectual honesty over impressive-looking results.

REFERENCES

- [1] H. Markowitz, "Portfolio Selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77-91, 1952.
- [2] W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, vol. 19, no. 3, pp. 425-442, 1964.
- [3] E. F. Fama and K. R. French, "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, vol. 47, no. 2, pp. 427-465, 1992.
- [4] D. H. Bailey, J. M. Borwein, M. Lopez de Prado, and Q. J. Zhu, "Pseudomathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance," *Notices of the AMS*, vol. 61, no. 5, pp. 458-471, 2014.
- [5] C. R. Harvey, Y. Liu, and H. Zhu, "...and the Cross-Section of Expected Returns," *The Review of Financial Studies*, vol. 29, no. 1, pp. 5-68, 2016.
- [6] M. Lopez de Prado, "The 10 Reasons Most Machine Learning Funds Fail," *The Journal of Portfolio Management*, vol. 44, no. 6, pp. 120-133, 2018.
- [7] A. Ang and G. Bekaert, "Regime Switches in Interest Rates," *Journal of Business & Economic Statistics*, vol. 20, no. 2, pp. 163-182, 2002.
- [8] M. Guidolin and A. Timmermann, "Asset Allocation under Multivariate Regime Switching," *Journal of Economic Dynamics and Control*, vol. 31, no. 11, pp. 3503-3544, 2007.
- [9] M. Kritzman, S. Page, and D. Turkington, "Regime Shifts: Implications for Dynamic Strategies," *Financial Analysts Journal*, vol. 68, no. 3, pp. 22-39, 2012.
- [10] P. Nystrup, H. Madsen, and E. Lindström, "Long Memory of Financial Time Series and Hidden Markov Models with Time-Varying Parameters," *Journal of Forecasting*, vol. 34, no. 8, pp. 701-712, 2015.