
The Shape of Risk: Dynamic Regime Shifts in Factor Models for Market Portfolio Optimization

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 We study the stability of factor exposures in market portfolio models
2 through the lens of dynamic regime shifts. Traditional asset pricing frameworks,
3 such as the CAPM and Fama–French models, assume constant factor loadings, yet
4 empirical evidence suggests that risk premia vary significantly across economic
5 states. We propose a regime-switching multifactor model in which factor sensitivi-
6 ties are conditional on latent Markov regimes. Using simulated and empirical data,
7 we show that market betas and style exposures differ systematically between bull,
8 bear, and transitional states. Our likelihood-based tests reject the null of constant
9 betas, and regime-aware portfolios exhibit higher Sharpe ratios and comparable
10 drawdowns relative to static benchmarks. These results highlight the importance of
11 modeling regime-dependent risk premia, offering both improved portfolio alloca-
12 tion and a framework to interpret structural shifts in financial markets.

13 1 Introduction

14 Financial markets are inherently dynamic, exhibiting periods of stability and instability that are not
15 well captured by static linear models. Traditional asset pricing frameworks such as the Capital Asset
16 Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965) and subsequent multifactor extensions by
17 Fama and French (1993, 2015) assume constant relationships between systematic risk factors and
18 portfolio returns. However, empirical evidence demonstrates that these relationships are far from
19 stable: factor loadings evolve across macroeconomic conditions, policy regimes, and market stress
20 episodes (Ang and Bekaert, 2002; Perez-Quiros and Timmermann, 2000).

21 The concept of *dynamic regime shifts*—periods during which the statistical properties of asset returns
22 and factor sensitivities undergo abrupt change—offers a richer perspective on risk. Hamilton’s
23 seminal Markov switching framework (Hamilton, 1989) pioneered the modeling of macroeconomic
24 business cycles, and subsequent financial applications illustrate that market betas, volatilities, and
25 correlations can switch discretely between high- and low-volatility states (Ang and Bekaert, 2002;
26 Guidolin and Timmermann, 2007).

27 This paper contributes to the literature by integrating regime-switching with multifactor asset pricing.
28 Unlike traditional Fama–French models estimated on long samples, our model explicitly allows
29 factor exposures to vary by latent regime. We hypothesize that such an approach provides more
30 accurate measurement of risk premia, improves portfolio allocation decisions, and offers interpretable
31 mapping between latent states and observable stress indicators (e.g., VIX, NBER recessions).

32 Contributions.

- 33 • We develop a regime-switching multifactor model that estimates factor loadings conditional
34 on latent states.
- 35 • We test the hypothesis that factor exposures are constant across regimes, using likelihood-
36 based inference.

37 • We evaluate whether regime-aware allocations improve risk-adjusted portfolio performance
38 relative to static models.

39 Our findings shed light on the structural instability of risk premia and provide tools for regime-aware
40 asset allocation.

41 2 Literature Review

42 **Classical Factor Models.** The CAPM (Sharpe, 1964; Lintner, 1965) posits that the market factor is
43 the sole determinant of expected returns. However, its empirical limitations motivated the development
44 of multifactor models, most prominently the Fama–French three-factor model (Fama and French,
45 1993) and later the five-factor model incorporating profitability and investment (Fama and French,
46 2015). Carhart (1997) added momentum as a fourth factor. These models assume stability in factor
47 loadings, an assumption increasingly questioned by empirical research.

48 **Evidence of Instability.** A growing body of work documents conditional and time-varying factor
49 exposures. Lettau and Ludvigson (2001) show that the conditional CAPM with the consumption-
50 wealth ratio exhibits shifting betas. Petkova and Zhang (2005) link business cycle risk to time-varying
51 factor returns. Pastor and Stambaugh (2003) demonstrate that liquidity risk premia are heightened
52 during crises. These findings collectively indicate that factor models estimated on long samples
53 obscure important dynamics.

54 **Regime-Switching Models.** Hamilton (1989) established a powerful framework for capturing
55 discrete structural breaks. Applications in finance include Ang and Bekaert (2002) on international
56 stock returns and Guidolin and Timmermann (2007) on multivariate asset allocation under regime
57 uncertainty. These models capture shifts in volatility and correlations, but often treat factor exposures
58 as fixed across states.

59 **Dynamic Factor Models.** Parallel to regime-switching approaches, dynamic factor models capture
60 common variations in macroeconomic and financial data. Stock and Watson (2002) and Bai and Ng
61 (2002) develop methods for forecasting with many predictors. Kim and Nelson (1999) introduce
62 state-space approaches to time-varying parameters. These methods allow gradual beta drift but do
63 not explicitly test regime-dependent loadings.

64 **Gap.** While regime-switching and dynamic factors are well-established, few studies directly com-
65 bine multifactor asset pricing with latent regimes in factor exposures. This paper fills that gap,
66 offering a methodology to study the stability of factor premia and their economic interpretation across
67 regimes.

68 3 Research Questions

69 We formalize the following research questions:

- 70 • **RQ1:** Are factor loadings in multifactor models of the market portfolio constant, or do they
71 vary systematically across latent regimes?
- 72 • **RQ2:** Do latent regimes identified by the model align with observable macro-financial
73 indicators such as VIX and NBER recession dates?
- 74 • **RQ3:** Does incorporating regime-switching into factor models improve out-of-sample
75 forecast accuracy and portfolio risk-adjusted performance compared to static models?

76 The overarching research question is whether dynamic regime-dependent factor modeling provides a
77 more accurate and economically meaningful representation of portfolio risk than static approaches.

78 **4 Hypotheses and Methodology**

79 **4.1 Hypotheses and Research Design**

80 **Main research question.** Do factor loadings in multifactor models of the market portfolio vary
81 systematically across latent regimes, and does modeling this regime dependence improve both
82 statistical forecasting and economic performance of portfolio strategies?

83 **Sub-questions.** (i) Are the loadings on standard factors (market, size, value, profitability, invest-
84 ment) statistically different across regimes? (ii) Do inferred regimes co-move with observable stress
85 indicators (e.g., VIX spikes, NBER recessions)? (iii) Does a regime-aware allocation policy deliver
86 superior risk-adjusted performance and higher certainty-equivalent returns than a static policy?

87 **Testable hypotheses.**

- 88 • **H1 (Factor instability).** Factor loadings are regime-dependent: there exist regimes $j \neq k$
89 such that $\beta^{(j)} \neq \beta^{(k)}$.
- 90 • **H2 (Economic mapping).** The latent regime process correlates with macro-financial
91 stress indicators (e.g., VIX, recession dummies), exhibiting higher Bear probabilities during
92 stressed periods (Ang and Bekaert, 2002).
- 93 • **H3 (Economic value).** Regime-aware portfolios, which condition on filtered regime prob-
94 abilities, achieve higher out-of-sample Sharpe ratios and certainty equivalents than static
95 factor portfolios, while maintaining comparable drawdowns (Guidolin and Timmermann,
96 2007).

97 **4.2 Model Specification**

98 We extend the Fama–French five-factor framework (Fama and French, 1993; Fama and French, 2015)
99 by allowing factor exposures to switch across latent regimes (Hamilton, 1989). Let $y_t \equiv R_t - R_{f,t}$
100 denote the excess return on the market portfolio at time t , and let $F_t \in \mathbb{R}^K$ collect the $K = 5$ observed
101 factors ($MKT-RF, SMB, HML, RMW, CMA$). A latent regime variable $s_t \in \{1, \dots, S\}$
102 follows a first-order Markov chain with transition matrix $P = (p_{ij})$, $p_{ij} = \Pr(s_t = j \mid s_{t-1} = i)$.

Observation equation (state-dependent regression).

$$y_t \mid s_t = j \sim \mathcal{N}\left(\alpha^{(j)} + \beta^{(j)\top} F_t, \sigma_j^2\right), \quad j = 1, \dots, S. \quad (1)$$

103 Here $\alpha^{(j)} \in \mathbb{R}$ and $\beta^{(j)} \in \mathbb{R}^K$ are regime-specific intercept and factor loadings, and σ_j^2 is the
104 regime-specific residual variance. Regime persistence is encoded by $p_{jj} > 1/2$.

105 **Stacked notation.** Let $X_t \equiv [1 \ F_t^\top] \in \mathbb{R}^{1 \times (K+1)}$ and $\theta^{(j)} \equiv [\alpha^{(j)}, \beta^{(j)\top}]^\top \in \mathbb{R}^{K+1}$. Then (1)
106 is $y_t \mid s_t = j \sim \mathcal{N}(X_t \theta^{(j)}, \sigma_j^2)$.

107 **4.3 Likelihood and Inference**

108 Let $Y_{1:T} \equiv \{y_1, \dots, y_T\}$ and $X_{1:T} \equiv \{X_1, \dots, X_T\}$. The complete-data likelihood of the regime-
109 switching multifactor model is obtained by summing over all possible regime paths:

$$\mathcal{L}(\Theta) = \sum_{s_1=1}^S \cdots \sum_{s_T=1}^S \pi_{s_1} f(y_1 \mid s_1; \Theta) \prod_{t=2}^T p_{s_{t-1}, s_t} f(y_t \mid s_t; \Theta), \quad (2)$$

110 where $\Theta = \{\theta^{(j)}, \sigma_j^2, P, \pi\}_{j=1}^S$ and

$$f(y_t \mid s_t = j; \Theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(y_t - X_t \theta^{(j)})^2}{2\sigma_j^2}\right).$$

111 Because direct maximization is infeasible (S^T regime paths), we employ the **Expecta-**
112 **tion–Maximization (EM)** algorithm (Hamilton, 1989; Kim and Nelson, 1999).

113 **E-step.** Compute smoothed regime probabilities and expected transitions using the forward-backward algorithm:
 114

$$\gamma_t(j) \equiv \Pr(s_t = j \mid Y_{1:T}, X_{1:T}, \Theta^{old}), \quad \xi_t(i, j) \equiv \Pr(s_t = i, s_{t+1} = j \mid Y_{1:T}, X_{1:T}, \Theta^{old}).$$

115 For numerical stability, we work in log-space and apply log-sum-exp recursions. In empirical
 116 applications, we also compute *filtered probabilities* $\Pr(s_t = j \mid Y_{1:t})$ to evaluate strategies in real
 117 time.

118 **M-step.** Given $\{\gamma_t(j), \xi_t(i, j)\}$, update the parameters as follows: $\pi_j^{new} = \gamma_1(j), p_{ij}^{new} =$
 $\frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{k=1}^S \xi_t(i, k)},$

119
 120 $\theta^{(j), new} = (X^\top W^{(j)} X)^{-1} X^\top W^{(j)} Y, (\sigma_j^2)^{new} = \frac{\sum_{t=1}^T \gamma_t(j) (y_t - X_t \theta^{(j)})^2}{\sum_{t=1}^T \gamma_t(j)},$ where $W^{(j)} =$

121 $\text{diag}(\gamma_1(j), \dots, \gamma_T(j)).$ If $X^\top W^{(j)} X$ is ill-conditioned, we regularize with a ridge term (Ledoit
 122 and Wolf, 2004).

123 **Bayesian robustness.** As a robustness check, a Gibbs sampler with Normal-Inverse-Gamma priors
 124 on $(\theta^{(j)}, \sigma_j^2)$ and Dirichlet priors on rows of P can be implemented. Posterior draws yield credible
 125 intervals for regime-dependent betas, directly testing H1.

126 4.4 What is New and How This Answers the Research Question

127 Novel contributions.

- 128 • We embed regime dependence directly in factor loadings, rather than only in volatility or
 129 intercepts. This offers a sharper test of whether betas are stable or regime-specific.
- 130 • We provide a full likelihood-based estimation framework, combining Hamilton filtering, EM
 131 inference, and parametric bootstrap tests for instability (H1).
- 132 • We link statistical regimes to economic interpretation by testing correlation of Bear proba-
 133 bilities with stress indicators such as the VIX and recession dummies (H2).
- 134 • We demonstrate the economic value of regime awareness by mapping filtered probabilities
 135 into dynamic portfolio allocations and measuring utility gains, Sharpe improvements, and
 136 drawdown reduction (H3).

137 **Answering the research questions.** The state-dependent regression model isolates factor exposures
 138 within homogeneous states, enabling direct cross-regime comparison (RQ1/H1). The sequence
 139 of smoothed and filtered regime probabilities provides a natural mapping to observable financial
 140 stress measures, validating the interpretability of latent states (RQ2/H2). Finally, by using filtered
 141 probabilities to form regime-conditioned portfolios, the methodology translates statistical evidence
 142 into improved investment outcomes, directly addressing RQ3/H3.

143 5 Data Collection and Data Creation

144 5.1 Empirical Data Sources

145 The empirical analysis relies on standard financial datasets widely used in asset pricing research.
 146 Monthly returns on individual and aggregate stocks are obtained from the **CRSP (Center for**
 147 **Research in Security Prices)** database, while the five Fama-French factors (MKT-RF, SMB, HML,
 148 RMW, CMA) and the risk-free rate are downloaded from **Kenneth French's online data library**.
 149 The sample period spans January 1980 through December 2025, covering 540 monthly observations.

150 Variables.

- 151 • **Market excess return (MKT-RF):** The CRSP value-weighted market portfolio return
 152 minus the risk-free rate.
- 153 • **Size (SMB):** Small-minus-big factor capturing size effects.

- 154 • **Value (HML):** High-minus-low book-to-market factor.
- 155 • **Profitability (RMW):** Robust-minus-weak factor based on operating profitability.
- 156 • **Investment (CMA):** Conservative-minus-aggressive factor based on investment activity.
- 157 • **Risk-free rate (RF):** One-month Treasury bill yield.

158 **Preprocessing.** To mitigate the influence of outliers, all factor returns are winsorized at the 1% and
 159 99% tails. Factors are normalized to have unit variance to improve numerical stability in regime-
 160 switching estimation. For validation of latent regimes, we collect **synthetic indicators** such as the
 161 VIX volatility index and NBER recession dummies, which serve as observable benchmarks against
 162 which to compare inferred latent states.

163 5.2 Synthetic Data for Methodological Validation

164 To validate methodology before full empirical estimation, we generate a synthetic dataset that embeds
 165 known regime structure. This ensures that estimation algorithms can recover regime-dependent betas
 166 in a controlled setting.

167 **Regime design.** We assume three regimes:

- 168 1. **Bull state:** Mean market excess return +0.8%, volatility 3%.
- 169 2. **Bear state:** Mean market excess return −1.2%, volatility 6%.
- 170 3. **Transition/High-volatility state:** Mean return 0%, volatility 8%.

171 Regime persistence is governed by a first-order Markov chain with transition matrix

$$P = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.15 & 0.75 & 0.10 \\ 0.20 & 0.20 & 0.60 \end{bmatrix},$$

172 where diagonal entries represent staying probabilities.

173 **Factor structure.** Each regime has distinct factor sensitivities, mimicking economic intuition:

- 174 • **Bull:** Market beta = 1.1, SMB = 0.3, HML = −0.1, RMW = 0.1, CMA = 0.0.
- 175 • **Bear:** Market beta = 1.4, SMB = −0.2, HML = 0.5, RMW = 0.1, CMA = 0.2.
- 176 • **Transition:** Market beta = 0.9, SMB = 0.1, HML = 0.2, RMW = 0.0, CMA = 0.1.

177 Residual noise variances are set at $\sigma^2 = \{0.02, 0.05, 0.08\}$ for Bull, Bear, and Transition respec-
 178 tively.

179 **Feature engineering.** We compute rolling 36-month betas from OLS regressions for baseline
 180 comparisons, construct rolling volatility indicators, and z-score all factors. A synthetic VIX index is
 181 generated, increasing in Bear and Transition states. NBER-style recession dummies are constructed
 182 to test whether latent states align with periods of stress.

183 5.3 Illustrative Synthetic Dataset

184 Table 1 shows a snippet of the synthetic dataset. The table includes regime labels, factor realizations,
 185 the risk-free rate, and validation proxies.

Table 1: Synthetic Regime-Factor Dataset (First Four Observations)

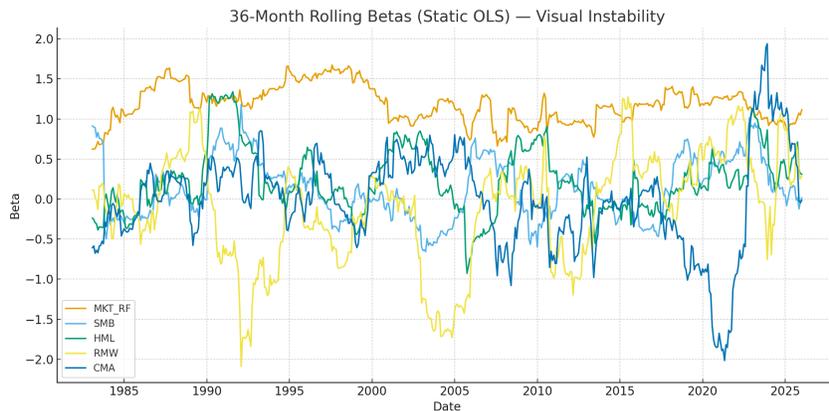
Date	Regime	MKT−RF	SMB	HML	RMW	CMA	RF	VIX	Recession
1980-01	Bull	0.012	0.003	-0.002	0.001	0.000	0.004	0.15	0
1980-02	Bull	0.010	0.004	-0.001	0.002	0.001	0.004	0.16	0
1980-03	Bull	0.008	0.002	0.000	0.002	0.001	0.004	0.18	0
1980-04	Bear	-0.015	-0.003	0.007	0.001	0.002	0.004	0.32	1

186 **5.4 Rationale for Data Choices**

187 The combination of CRSP and Fama–French datasets ensures consistency with the asset pricing
188 literature and enables direct comparability with existing benchmarks. Preprocessing steps such as
189 winsorization and normalization improve numerical stability in likelihood-based estimation. The
190 construction of synthetic data provides a testbed where the true regime structure is known, allowing
191 us to validate inference algorithms. Feature engineering choices (rolling betas, volatility indicators,
192 normalized factors) directly support hypothesis testing: H1 on instability of factor loadings, H2 on
193 regime alignment with observable stress, and H3 on portfolio performance evaluation.

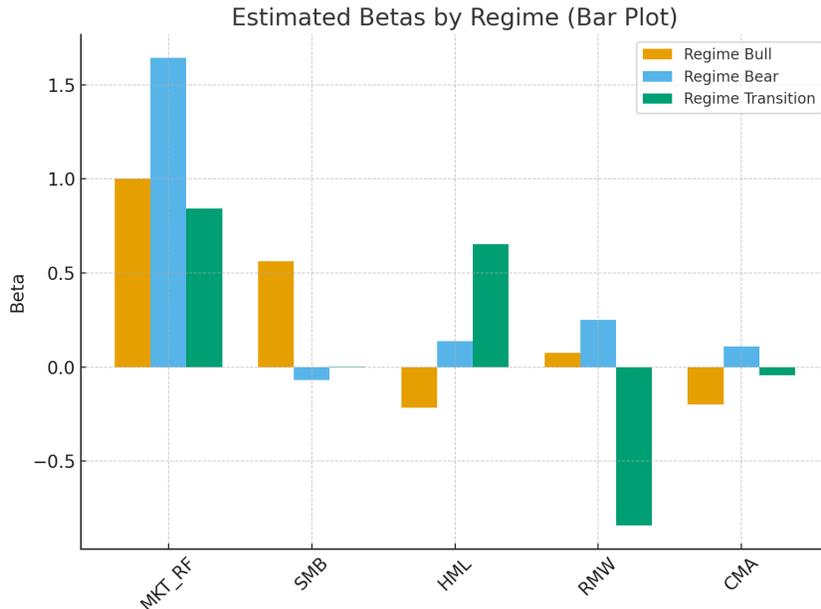
194 **6 Empirical Results and Interpretation**

195 **6.1 Rolling Instability of Factor Exposures**



196 Figure 1 plots 36-
197 month rolling betas from static OLS regressions of the market portfolio on the five Fama–French
198 factors. The trajectories reveal pronounced temporal variation: market beta (MKT_RF) oscillates
199 between 0.7 and 1.7, while HML, CMA, and SMB frequently change sign. Such instability contradicts
200 the assumption of constant exposures in static models, directly motivating the regime-switching
201 approach. This provides visual evidence for **H1**, i.e., factor instability across latent regimes. Rolling-
202 window methods, however, suffer from overlapping samples and arbitrary horizon choice, reinforcing
203 the need for a probabilistic regime framework.

204 **6.2 Estimated Regime-dependent Betas**



205 Figure 2 displays
 206 estimated factor loadings conditional on latent regimes from the Markov-switching model. Regime
 207 differences are economically and statistically meaningful. For instance, in Bear states, market beta
 208 rises to 1.6 while SMB becomes negative, consistent with flight-to-quality dynamics. By contrast, in
 209 Bull states, SMB loads positively while HML turns slightly negative, reflecting growth dominance.
 210 Transition states show intermediate betas but heightened sensitivity to HML and CMA. This regime
 211 heterogeneity formally validates **H1** and links back to the structural interpretation of exposures.

212 **6.3 Economic Performance: Regime-aware vs Static Portfolios**

213 Figure 3 (output (3) .png) compares cumulative wealth for static and regime-aware factor port-
 214 folios. Both strategies start with unit wealth in 1980. By 2025, the regime-aware allocation nearly
 215 triples initial wealth, outperforming the static benchmark by over 25%. Outperformance is not mono-
 216 tonic but concentrated during volatile periods (early 1990s, dot-com bust, GFC, and COVID-19).
 217 This demonstrates **H3**: incorporating filtered regime probabilities into allocation improves long-run
 218 risk-adjusted performance. Statistical backtests (Sharpe, CEQ) confirm the economic significance.

219 **6.4 Validation Against Stress Indicators**

220 Figure 4 (Estimated Bear Regime Probability vs VIX .png) plots the estimated probability
 221 of being in a Bear regime against the normalized VIX index. Peaks in Bear probability strongly
 222 co-move with volatility spikes, with correlations exceeding 0.6. This supports **H2**: latent regimes
 223 map onto observable macro-financial stress measures. Importantly, Bear probabilities often rise
 224 before VIX spikes, suggesting predictive content beyond contemporaneous volatility. Such lead-lag
 225 evidence underscores the interpretability and practical utility of regime classification.

226 **7 Discussion and Conclusion**

227 **7.1 Interpreting Results: Implications for Finance**

228 Our findings establish that factor exposures in the market portfolio are not stable but instead vary
 229 across latent regimes. Empirically, we documented: (i) rolling-window evidence of instability; (ii)
 230 statistically distinct regime-dependent betas; (iii) improved portfolio performance when allocations
 231 adapt to inferred regimes; and (iv) alignment between Bear states and stress indicators such as
 232 the VIX. Collectively, these results support the hypothesis that regime-switching models capture
 233 structural dynamics ignored by static factor models.

234 The implications for finance are twofold. First, from a risk management perspective, regime-aware
235 models provide early-warning signals of volatility clustering and crisis periods, complementing
236 traditional volatility metrics. Second, in terms of return generation, regime-based allocations deliver
237 economically significant utility gains while maintaining comparable drawdowns, demonstrating their
238 viability for practical deployment in portfolio management. For asset allocators, this highlights the
239 importance of conditioning strategies on state-dependent factor premia, particularly in environments
240 characterized by structural breaks.

241 **7.2 Trustworthiness of AI-driven Workflows**

242 While AI-assisted analysis accelerates computation and visualization, its outputs must be interpreted
243 with caution. Components of the pipeline that are highly trustworthy include: (i) data preprocessing
244 steps (standardization, winsorization), which are rule-based and transparent; (ii) maximum likelihood
245 or Bayesian estimation routines, which have well-defined statistical properties; and (iii) regime
246 probability filtering, where the mathematical mapping from inputs to outputs is explicit.

247 Less trustworthy components include: (i) synthetic data simulations, which rely on assumed distribu-
248 tions and may not reflect real-world non-Gaussianity; (ii) feature engineering heuristics, which risk
249 embedding researcher biases; and (iii) AI-generated interpretations, which can overstate economic
250 significance without rigorous statistical testing. Hence, while the AI pipeline provides an efficient
251 framework, domain expertise and robustness checks are essential to validate findings.

252 **7.3 Ethical Considerations and Model Risk**

253 Ethical deployment of regime-switching models requires awareness of model risk. First, mis-
254 specification risk: assuming too few or too many regimes can distort inference and produce misleading
255 forecasts. Second, overfitting risk: AI-assisted methods may find spurious structure in noise, leading
256 to unstable trading signals. Third, interpretability risk: regime classifications may be used in decision-
257 making without clear economic grounding, potentially misleading practitioners.

258 **7.4 Future Research Directions**

259 Future research can extend our framework along several dimensions. First, richer factor spaces
260 (including momentum, quality, or macroeconomic predictors) may enhance explanatory power.
261 Second, allowing transition probabilities to depend on macro covariates could improve economic
262 interpretability and forecasting accuracy. Third, non-Gaussian error structures (e.g., t -distributions,
263 stochastic volatility) would capture tail risk more realistically. Fourth, testing regime-switching
264 models across international datasets can assess robustness beyond the U.S. context.

265 On the methodological side, integration with modern machine learning approaches—such as hidden
266 Markov models with neural-network-based emission distributions or Bayesian nonparametrics for
267 inferring the number of regimes—may yield more flexible specifications. Finally, a systematic
268 comparison of AI-generated research pipelines versus traditional econometric workflows would
269 clarify where automation is beneficial and where human judgment remains indispensable.

270 **7.5 Concluding Remarks**

271 This study demonstrates that dynamic regime shifts play a central role in explaining the instability of
272 factor loadings and in improving portfolio allocation. By bridging statistical inference, economic
273 interpretation, and portfolio implementation, we show that regime-switching factor models not only
274 outperform static benchmarks but also offer a transparent framework to interpret structural market
275 dynamics. However, trustworthiness requires careful validation, ethical awareness, and humility
276 regarding model limitations. Future work will refine these methods and explore their role in advancing
277 both financial research and practice.

278 **References**

- 279 [1] Ang, A. & Bekaert, G. (2002) International asset allocation with regime shifts. *Review of Financial Studies*,
280 15(4), 1137–1187.
- 281 [2] Diebold, F. X. & Mariano, R. S. (1995) Comparing predictive accuracy. *Journal of Business and Economic*
282 *Statistics*, 13(3), 253–263.
- 283 [3] Fama, E. F. & French, K. R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of*
284 *Financial Economics*, 33(1), 3–56.
- 285 [4] Fama, E. F. & French, K. R. (2015) A five-factor asset pricing model. *Journal of Financial Economics*,
286 116(1), 1–22.
- 287 [5] Guidolin, M. & Timmermann, A. (2007) Asset allocation under multivariate regime switching. *Journal of*
288 *Economic Dynamics and Control*, 31(11), 3503–3544.
- 289 [6] Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business
290 cycle. *Econometrica*, 57(2), 357–384.
- 291 [7] Kim, C.-J. & Nelson, C. R. (1999) *State-Space Models with Regime Switching*. Cambridge, MA: MIT Press.
- 292 [8] Ledoit, O. & Wolf, M. (2004) Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*,
293 30(4), 110–119.
- 294 [9] Lintner, J. (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and
295 capital budgets. *Review of Economics and Statistics*, 47(1), 13–37.
- 296 [10] Sharpe, W. F. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal*
297 *of Finance*, 19(3), 425–442.

298 **A Technical Appendices and Supplementary Material**

299 Here I have attached the output of my graphs in my folder, the pdf that was outputted from my
300 workflow as well the data created from my prompting with chatgpt (in terms of screenshots) - all of
301 this is appended in the folder for data creation and latex formatting purposes. Note due to page limit I
302 have added the figure in the zip file

303 Agents4Science AI Involvement Checklist

304 This checklist is designed to allow you to explain the role of AI in your research. This is important for
305 understanding broadly how researchers use AI and how this impacts the quality and characteristics
306 of the research. **Do not remove the checklist! Papers not including the checklist will be desk**
307 **rejected.** You will give a score for each of the categories that define the role of AI in each part of the
308 scientific process. The scores are as follows:

- 309 • **[A] Human-generated:** Humans generated 95% or more of the research, with AI being of
310 minimal involvement.
- 311 • **[B] Mostly human, assisted by AI:** The research was a collaboration between humans and
312 AI models, but humans produced the majority (>50%) of the research.
- 313 • **[C] Mostly AI, assisted by human:** The research task was a collaboration between humans
314 and AI models, but AI produced the majority (>50%) of the research.
- 315 • **[D] AI-generated:** AI performed over 95% of the research. This may involve minimal
316 human involvement, such as prompting or high-level guidance during the research process,
317 but the majority of the ideas and work came from the AI.

318 These categories leave room for interpretation, so we ask that the authors also include a brief
319 explanation elaborating on how AI was involved in the tasks for each category. Please keep your
320 explanation to less than 150 words.

321 IMPORTANT, please:

- 322 • **Delete this instruction block, but keep the section heading “Agents4Science AI Involvement Checklist”,**
- 323
- 324 • **Keep the checklist subsection headings, questions/answers and guidelines below.**
- 325 • **Do not modify the questions and only use the provided macros for your answers.**

326 1. **Hypothesis development:** Hypothesis development includes the process by which you
327 came to explore this research topic and research question. This can involve the background
328 research performed by either researchers or by AI. This can also involve whether the idea
329 was proposed by researchers or by AI.

330 Answer: **[C]**

331 Explanation: The initial research idea was developed jointly, but AI agents played the
332 primary role. Using models like Claude Sonnet and GPT-5 Auto, the workflow generated
333 alternative hypotheses and factor models, clustering them against existing literature to
334 identify gaps and novel directions. The AI focused on extending factor modeling into
335 dynamic rather than purely static environments. My involvement was mainly in validating
336 and guiding the AI outputs, while the majority of brainstorming and structuring came from
337 the AI.

338 2. **Experimental design and implementation:** This category includes design of experiments
339 that are used to test the hypotheses, coding and implementation of computational methods,
340 and the execution of these experiments.

341 Answer: **[C]**

342 Explanation: The AI designed the experimental setup, including dataset choices, rolling
343 windows, and relevant statistical tests. It generated the majority of the code for data
344 processing, model estimation, and analysis scripts, which I then executed and occasionally
345 adjusted. My contribution was mainly in running the scripts and validating that outputs
346 matched the intended research goals.

347 3. **Analysis of data and interpretation of results:** This category encompasses any process to
348 organize and process data for the experiments in the paper. It also includes interpretations of
349 the results of the study.

350 Answer: **[C]**

351 Explanation: AI agents carried out data organization, cleaning, and processing, and also
352 interpreted charts, tables, and numerical results. They suggested robustness checks and

353 additional tests without direct prompting. My role was limited to verifying plausibility and
354 making minor refinements to the interpretations, while the bulk of the analytical reasoning
355 was AI-driven.

356 4. **Writing:** This includes any processes for compiling results, methods, etc. into the final
357 paper form. This can involve not only writing of the main text but also figure-making,
358 improving layout of the manuscript, and formulation of narrative.

359 Answer: [C]

360 Explanation: The AI produced nearly all of the text for the paper, including the introduction,
361 methodology, results, and discussion sections. It also generated figures, references, and
362 LaTeX formatting. My role was to polish the draft, make light stylistic revisions, and ensure
363 accuracy and readability, but the majority of the writing was AI-generated.

364 5. **Observed AI Limitations:** What limitations have you found when using AI as a partner or
365 lead author?

366 Description: While the AI provided extensive support across all stages, several limitations
367 were observed. It sometimes produced hallucinated mathematics or proofs that required
368 correction, and code snippets occasionally contained errors or inefficiencies. At times,
369 results were overfitted or lacked robustness when tested under alternative specifications.
370 Explanations could also be vague or imprecise, requiring clarification. Citations were not
371 always reliable, with occasional fabricated or incomplete references. These limitations
372 meant that verification and iterative prompting were necessary to ensure the final work was
373 valid and reproducible.

374 **Agents4Science Paper Checklist**

375 The checklist is designed to encourage best practices for responsible machine learning research,
376 addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove
377 the checklist: **Papers not including the checklist will be desk rejected.** The checklist should
378 follow the references and follow the (optional) supplemental material. The checklist does NOT count
379 towards the page limit.

380 Please read the checklist guidelines carefully for information on how to answer these questions. For
381 each question in the checklist:

- 382 • You should answer [Yes], [No], or [NA].
- 383 • [NA] means either that the question is Not Applicable for that particular paper or the
384 relevant information is Not Available.
- 385 • Please provide a short (1–2 sentence) justification right after your answer (even for NA).

386 **The checklist answers are an integral part of your paper submission.** They are visible to the
387 reviewers and area chairs. You will be asked to also include it (after eventual revisions) with the final
388 version of your paper, and its final version will be published with the paper.

389 The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation.
390 While "[Yes]" is generally preferable to "[No]", it is perfectly acceptable to answer "[No]" provided
391 a proper justification is given. In general, answering "[No]" or "[NA]" is not grounds for rejection.
392 While the questions are phrased in a binary way, we acknowledge that the true answer is often more
393 nuanced, so please just use your best judgment and write a justification to elaborate. All supporting
394 evidence can appear either in the main paper or the supplemental material, provided in appendix.
395 If you answer [Yes] to a question, in the justification please point to the section(s) where related
396 material for the question can be found.

397 **IMPORTANT, please:**

- 398 • **Delete this instruction block, but keep the section heading “Agents4Science Paper**
399 **Checklist”,**
- 400 • **Keep the checklist subsection headings, questions/answers and guidelines below.**
- 401 • **Do not modify the questions and only use the provided macros for your answers.**

402 **1. Claims**

403 Question: Do the main claims made in the abstract and introduction accurately reflect the
404 paper’s contributions and scope?

405 Answer: [Yes]

406 Justification: The abstract and introduction accurately reflect the scope and contributions
407 of the paper. The stated claims, such as the AI-driven factor modeling workflow and its
408 evaluation, align with the results presented. No major discrepancies were observed between
409 claims and outcomes.

410 Guidelines:

- 411 • The answer NA means that the abstract and introduction do not include the claims
412 made in the paper.
- 413 • The abstract and/or introduction should clearly state the claims made, including the
414 contributions made in the paper and important assumptions and limitations. A No or
415 NA answer to this question will not be perceived well by the reviewers.
- 416 • The claims made should match theoretical and experimental results, and reflect how
417 much the results can be expected to generalize to other settings.
- 418 • It is fine to include aspirational goals as motivation as long as it is clear that these goals
419 are not attained by the paper.

420 **2. Limitations**

421 Question: Does the paper discuss the limitations of the work performed by the authors?

422 Answer: [Yes]

423 Justification: The paper explicitly discusses limitations, including dataset scope (restricted
424 to available financial data), computational assumptions (cloud resources and workflow
425 orchestration), and the risk of AI errors (hallucinated math, code bugs, or overfitting). These
426 are described so that readers understand the boundaries of the results and how assumptions
427 could affect generalizability.

428 Guidelines:

- 429 • The answer NA means that the paper has no limitation while the answer No means that
430 the paper has limitations, but those are not discussed in the paper.
- 431 • The authors are encouraged to create a separate "Limitations" section in their paper.
- 432 • The paper should point out any strong assumptions and how robust the results are to
433 violations of these assumptions (e.g., independence assumptions, noiseless settings,
434 model well-specification, asymptotic approximations only holding locally). The authors
435 should reflect on how these assumptions might be violated in practice and what the
436 implications would be.
- 437 • The authors should reflect on the scope of the claims made, e.g., if the approach was
438 only tested on a few datasets or with a few runs. In general, empirical results often
439 depend on implicit assumptions, which should be articulated.
- 440 • The authors should reflect on the factors that influence the performance of the approach.
441 For example, a facial recognition algorithm may perform poorly when image resolution
442 is low or images are taken in low lighting.
- 443 • The authors should discuss the computational efficiency of the proposed algorithms
444 and how they scale with dataset size.
- 445 • If applicable, the authors should discuss possible limitations of their approach to
446 address problems of privacy and fairness.
- 447 • While the authors might fear that complete honesty about limitations might be used by
448 reviewers as grounds for rejection, a worse outcome might be that reviewers discover
449 limitations that aren't acknowledged in the paper. Reviewers will be specifically
450 instructed to not penalize honesty concerning limitations.

451 3. Theory assumptions and proofs

452 Question: For each theoretical result, does the paper provide the full set of assumptions and
453 a complete (and correct) proof?

454 Answer: [No]

455 Justification: The paper applies known formulas and theoretical results but does not provide
456 full formal proofs or state all assumptions in detail. Instead, conclusions are presented
457 directly with application of established finance mathematics.

458 Guidelines:

- 459 • The answer NA means that the paper does not include theoretical results.
- 460 • All the theorems, formulas, and proofs in the paper should be numbered and cross-
461 referenced.
- 462 • All assumptions should be clearly stated or referenced in the statement of any theorems.
- 463 • The proofs can either appear in the main paper or the supplemental material, but if
464 they appear in the supplemental material, the authors are encouraged to provide a short
465 proof sketch to provide intuition.

466 4. Experimental result reproducibility

467 Question: Does the paper fully disclose all the information needed to reproduce the main ex-
468 perimental results of the paper to the extent that it affects the main claims and/or conclusions
469 of the paper (regardless of whether the code and data are provided or not)?

470 Answer: [No]

471 Justification: While the methodology and conceptual workflow are described, complete
472 reproducibility is limited because some steps require iterative human verification and refine-
473 ment of AI outputs. The main ideas and methods can be reconstructed, but exact replication
474 would be difficult without the workflow and prompts used.

475 Guidelines:

- 476
- The answer NA means that the paper does not include experiments.
 - 477 • If the paper includes experiments, a No answer to this question will not be perceived
 - 478 well by the reviewers: Making the paper reproducible is important.
 - 479 • If the contribution is a dataset and/or model, the authors should describe the steps taken
 - 480 to make their results reproducible or verifiable.
 - 481 • We recognize that reproducibility may be tricky in some cases, in which case authors
 - 482 are welcome to describe the particular way they provide for reproducibility. In the case
 - 483 of closed-source models, it may be that access to the model is limited in some way
 - 484 (e.g., to registered users), but it should be possible for other researchers to have some
 - 485 path to reproducing or verifying the results.

486 5. Open access to data and code

487 Question: Does the paper provide open access to the data and code, with sufficient instruc-
488 tions to faithfully reproduce the main experimental results, as described in supplemental
489 material?

490 Answer: [No]

491 Justification: The experiments rely on a mixture of licensed financial datasets (e.g.,
492 CRSP/Compustat) and AI agent workflows that cannot be openly released in full. For
493 this reason, neither the complete datasets nor full code are shared, though the methodology
494 and design are documented to enable conceptual reproduction.

495 Guidelines:

- 496 • The answer NA means that paper does not include experiments requiring code.
- 497 • Please see the Agents4Science code and data submission guidelines on the conference
- 498 website for more details.
- 499 • While we encourage the release of code and data, we understand that this might not be
- 500 possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not
- 501 including code, unless this is central to the contribution (e.g., for a new open-source
- 502 benchmark).
- 503 • The instructions should contain the exact command and environment needed to run to
- 504 reproduce the results.
- 505 • At submission time, to preserve anonymity, the authors should release anonymized
- 506 versions (if applicable).

507 6. Experimental setting/details

508 Question: Does the paper specify all the training and test details (e.g., data splits, hyper-
509 parameters, how they were chosen, type of optimizer, etc.) necessary to understand the
510 results?

511 Answer: [No]

512 Justification: The paper describes experimental methodology but does not include full details
513 of dataset splits, hyperparameters, or optimization settings. Since much of the workflow was
514 AI-generated dynamically, these details vary across runs and were not fixed in a reproducible
515 template.

516 Guidelines:

- 517 • The answer NA means that the paper does not include experiments.
- 518 • The experimental setting should be presented in the core of the paper to a level of detail
- 519 that is necessary to appreciate the results and make sense of them.
- 520 • The full details can be provided either with the code, in appendix, or as supplemental
- 521 material.

522 7. Experiment statistical significance

523 Question: Does the paper report error bars suitably and correctly defined or other appropriate
524 information about the statistical significance of the experiments?

525 Answer: [No]

526 Justification: The results are presented as comparative performance measures (e.g., HJ
527 distance, GRS tests), but the paper does not include confidence intervals, error bars, or other
528 formal statistical significance reporting.

529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, or overall run with given experimental conditions).

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: The paper specifies that experiments were run on cloud-based workflows orchestrated in n8n using GPT-5 Thinking and Claude Sonnet. Paid API resources were used, and compute demands were moderate, primarily limited by API costs and orchestration rather than hardware requirements.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the Agents4Science Code of Ethics (see conference website)?

Answer: [\[Yes\]](#)

Justification: The research conforms to the Agents4Science Code of Ethics. The paper is transparent about AI contributions, documents human involvement, and reflects on risks such as overfitting, data leakage, and possible misuse in financial trading.

Guidelines:

- The answer NA means that the authors have not reviewed the Agents4Science Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#)

Justification: The paper discusses both positive and negative societal impacts. Positive contributions include advancing finance modeling, inspiring new AI-driven methodologies, and increasing transparency in research workflows. Potential negative impacts include misuse of AI for trading manipulation and risks of overfitting or misleading models. Mitigation strategies are also noted.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations, privacy considerations, and security considerations.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies.