Regime-Aware Risk and Portfolio Allocation in the S&P 500 (H1 2025)

Ryan Luo and Peter Quawas
Department of Electrical and Computer Engineering
University of California San Diego

***Abstract*—This project analyzes daily trade data for S&P 500 companies during the first half of 2025 to study how risk and return vary across time and sectors. We construct an equal weight pseudo index from individual stock returns and apply time series diagnostics, GARCH volatility modeling, and a two state Markov switching model to identify high and low volatility regimes. Sector tail risk is measured using historical Value at Risk (VaR) and Expected Shortfall (ES). We compare equal weight, inverse volatility, and minimum variance portfolios using a shrinkage covariance estimator. The results show clear volatility clustering and distinct regimes, with high volatility periods aligned with drawdowns. Portfolios that respond to risk achieve shallower drawdowns and better risk adjusted performance, which illustrates how a regime based allocation can support more robust investment decisions.**

I. Introduction

Equity market volatility is not constant. Calm periods are often interrupted by episodes of stress that can cause concentrated portfolio losses. Standard allocations that ignore time varying risk may be suboptimal when volatility regimes differ sharply. This project addresses three questions for January through June 2025:

1) How strong are volatility regimes in the S&P 500?
2) How does tail risk differ across sectors?
3) Can simple allocation rules based on risk, such as inverse volatility and minimum variance portfolios, improve risk adjusted performance relative to equal weight portfolios?

Understanding volatility regimes and sector tail risk is important for position sizing, risk budgeting, and capital allocation. GARCH type models and Markov switching regimes are standard tools for capturing volatility clustering and regime changes [1], [2]. Combining these with portfolio optimization and covariance shrinkage [3] gives a practical example of how statistically informed risk models can support portfolio construction in noisy, finite sample settings.

II. Data Description and Preprocessing

A. Dataset

We use the Kaggle dataset "S&P 500 Stocks Trade Data for First 6 Months of 2025," which contains daily OHLCV fields (open, high, low, close, adjusted close, volume) for S&P 500 constituents from early January through June 30, 2025. Each row is a ticker day observation. Some days such as holidays are missing uniformly, but the panel is approximately balanced over this horizon.

B. Preprocessing Steps

We pivot the data into a wide date by ticker matrix of adjusted close prices and compute log returns

$$r_{t,i} = \log P_{t,i} - \log P_{t-1,i}, \tag{1}$$

for stock i on day t. The first observation per ticker is dropped, which yields a balanced panel of daily returns. An equal weight pseudo index is the cross sectional average of ticker returns each day.

Additional preprocessing includes alignment of trading days and removal of rows with incomplete cross sectional coverage, winsorization or removal of obvious data errors such as extreme returns beyond realistic market moves, and mapping each ticker to a sector using a public S&P 500 constituents file, with "Unknown" assigned when the mapping is unavailable. The result is a clean panel of daily log returns with sector labels.

III. Methods

A. Time Series Diagnostics

We analyze the equal weight index returns using a histogram with an overlaid Normal density and a QQ plot to assess deviations from normality. Autocorrelation functions of returns and squared returns reveal serial dependence and volatility clustering. As is typical for equity indexes, raw return autocorrelation is weak, while squared returns show strong positive autocorrelation, which motivates conditional heteroskedasticity models.

B. GARCH(1,1) with Student-t Innovations

We fit a GARCH(1,1) model with Student-t innovations using the `arch` Python library [4]:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \tag{2}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{3}$$

where z_t follows a standardized Student-t distribution. The conditional variance depends on one lag of the squared shock and one lag of previous variance. The Student-t distribution accommodates heavy tailed shocks better than Gaussian innovations. We examine the persistence $\alpha_1 + \beta_1$ and the conditional volatility series $\hat{\sigma}_t$.

C. Markov Switching Volatility Regimes

To identify discrete volatility regimes, we estimate a two state Markov switching model [1] using `statsmodels` [5]. Let $S_t \in \{1, 2\}$ denote the latent regime:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t \mid S_t = s \sim \mathcal{N}(0, \sigma_s^2). \tag{4}$$

The regime evolves as a first order Markov chain with transition probabilities $P_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$. Maximum likelihood yields regime specific variances and transition probabilities. Smoothed probabilities $\Pr(S_t = \text{high vol} \mid \text{data})$ define a time series of regime labels, which we overlay on the pseudo index level to compare portfolio performance across calm and turbulent periods.

D. Sector Tail Risk: VaR and ES

For sector level risk, daily sector returns are cross sectional averages of constituent returns. For sector k, we compute historical 95% Value at Risk and Expected Shortfall:

$$\mathrm{VaR}^k_{0.95} = q_{0.05}(r^k), \tag{5}$$

$$\mathrm{ES}^k_{0.95} = \mathbb{E}[r^k \mid r^k \le q_{0.05}(r^k)], \tag{6}$$

where $q_{0.05}(r^k)$ is the 5th percentile of sector k's return distribution. The empirical breach rate, defined as the fraction of days with $r^k_t \le \widehat{\mathrm{VaR}}^k_{0.95}$, serves as a simple VaR backtest.

E. Portfolio Construction

Using the ticker by day returns matrix, we form three daily rebalanced long only portfolios:

- **Equal weight (EW):** $w_i = 1/N$ for N stocks.
- **Inverse volatility (IV):** $w_i \propto 1/\hat{\sigma}_i$, based on each stock's return standard deviation.
- **Minimum variance (MV):** weights that minimize $w^\top \Sigma w$ subject to $w^\top \mathbf{1} = 1$, $w \ge 0$, where Σ is a Ledoit Wolf shrinkage covariance estimate [3].

For each portfolio, we compute daily returns, annualized return and volatility, a Sharpe ratio that assumes a zero risk free rate, and maximum drawdown. Metrics are also computed conditional on high and low volatility regimes.

IV. Results and Visualization

A. Diagnostics and Volatility Clustering

Fig. 1 displays the equal weight index histogram, QQ plot, and autocorrelation functions of returns and squared returns. The empirical distribution has heavier tails than a Normal fit, and the autocorrelation of squared returns is significantly positive over multiple lags. These patterns confirm volatility clustering and support the use of conditional volatility models.

B. GARCH Results

The GARCH(1,1) fit yields persistent conditional variance with $\alpha_1 + \beta_1$ close to but below one, which indicates high volatility persistence. The Student-t degrees of freedom parameter is finite, which captures heavy tails. Fig. 2 plots returns alongside GARCH conditional volatility, and volatility spikes line up with visually turbulent periods.

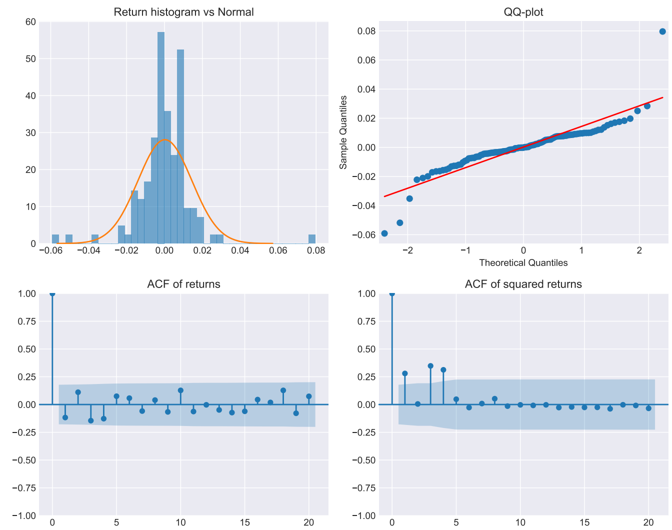

Fig. 1. Equal weight index diagnostics: histogram versus Normal fit (top left), Q Q plot (top right), autocorrelation of returns (bottom left), and autocorrelation of squared returns (bottom right). Heavy tails and positive autocorrelation in squared returns are evident.

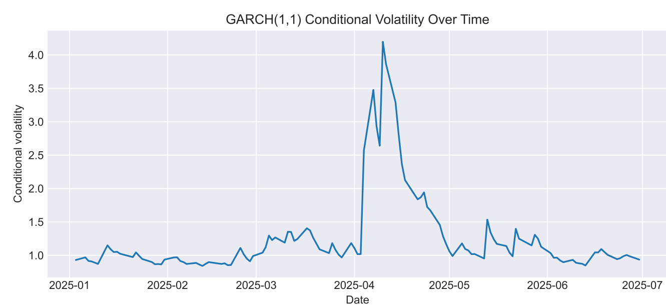

Fig. 2. Student-t GARCH(1,1) conditional volatility (shaded) for the equal weight index. Spikes correspond to periods of elevated market stress.

C. Markov Switching Regimes

Fig. 3 overlays the pseudo index level with shaded high volatility regimes and shows smoothed regime probabilities. The two regimes differ substantially in variance, and high volatility windows coincide with drawdowns. This shows that a discrete regime framework can capture meaningful shifts in market risk.

D. Sector VaR and ES

Fig. 4 reports historical 95% VaR and ES by sector. Over H1 2025, the aggregate sector has a VaR between 1% and 2% daily loss and an ES around 3% in the worst 5% of days, with breach frequencies close to the nominal 5%. More cyclical sectors show more negative VaR and ES, which indicates larger expected losses on extreme down days.

E. Portfolio Performance

Table I summarizes performance metrics for the three portfolios. Inverse volatility delivers higher annualized return and lower volatility than equal weight, which improves the return

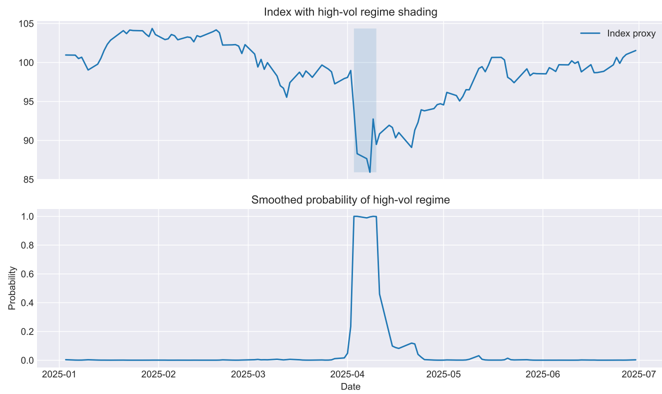

Fig. 3. Pseudo index level with high volatility regimes shaded (top) and smoothed probability of the high volatility regime (bottom). Regime switches align with drawdown periods.

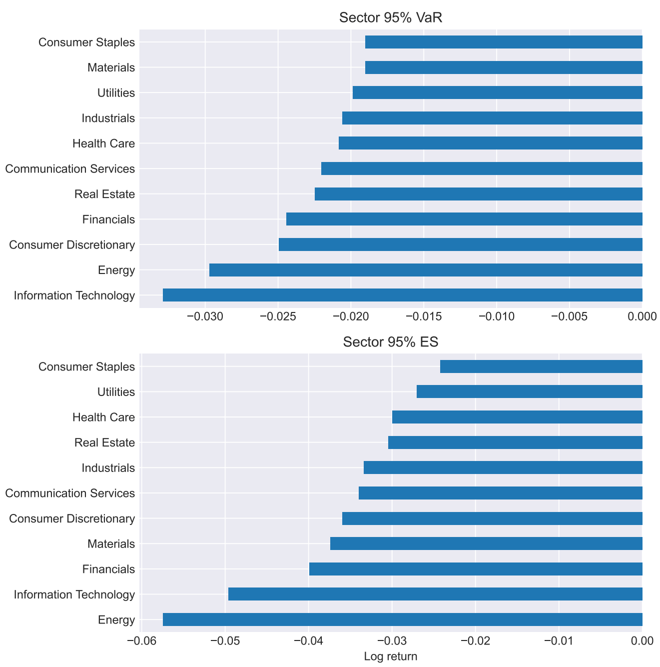

Fig. 4. Historical 95% VaR and ES by sector. Cyclical sectors show greater tail risk than defensive sectors.

to risk tradeoff. The minimum variance portfolio is the most defensive, with the lowest volatility and intermediate returns. Both IV and MV have smaller maximum drawdowns than EW.

Fig. 5 plots cumulative returns and drawdowns for the three strategies with regime shading. The risk sensitive portfolios keep better downside protection when volatility is elevated, which supports the value of allocations that respond to risk during stress periods.

V. Implications and Conclusion

A. Key Findings

The analysis shows that S&P 500 risk over H1 2025 is highly time varying, with distinct high and low volatility regimes. Heavy tails and volatility clustering break simple i.i.d.

TABLE I
PORTFOLIO METRICS (JAN–JUN 2025)

Portfolio	Ann. Ret.	Ann. Vol.	Sharpe	Max DD
EW	1.2%	22.5%	0.05	−17.7%
IV	3.5%	19.9%	0.17	−14.5%
MV	0.9%	18.7%	0.05	−15.6%

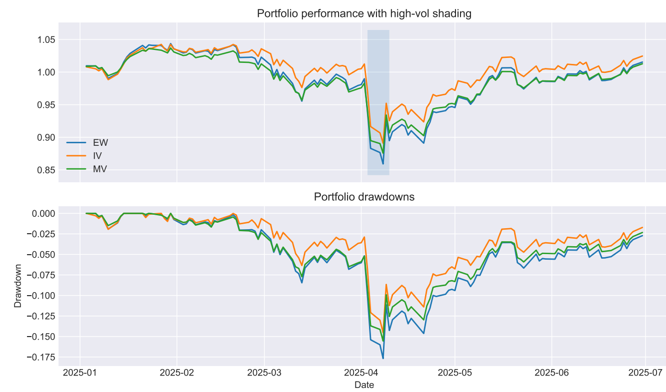

Fig. 5. Cumulative returns (top) and drawdowns (bottom) for EW, IV, and MV portfolios. Shaded regions indicate high volatility regimes. IV and MV show smaller drawdowns during turbulent periods.

assumptions and support the use of conditional volatility and regime models. From a portfolio perspective, inverse volatility and minimum variance strategies with covariance shrinkage reach more favorable risk adjusted performance and smaller drawdowns than a naive equal weight allocation, especially when volatility is high. Sector level VaR and ES show that tail risk is concentrated in certain sectors, which can guide risk budgeting and sector tilts.

B. Practical Implications

For practitioners, these results suggest that including volatility regime information and covariance shrinkage in portfolio construction can improve robustness without excessively complex optimization. Simple rules such as inverse volatility weighting already give meaningful drawdown reduction. Monitoring regime probabilities can also inform dynamic position sizing and hedging decisions.

C. Limitations and Future Work

Limitations include the short six month horizon of our dataset, construction of an equal weight index rather than a cap weighted index, omission of transaction costs and turnover constraints, and reliance on a two state regime specification. Future work could extend the horizon, add macro predictors to regime probabilities, explore alternative volatility models such as EGARCH or GJR GARCH, and implement explicit regime based allocation rules that scale exposure based on inferred risk states.

Appendix: Code

A complete Python implementation is provided in the accompanying Jupyter notebook, which loads the Kaggle dataset, performs preprocessing, estimates GARCH and Markov switching models using the `arch` [4] and `statsmodels` [5] libraries, computes sector VaR and ES, and constructs the three portfolios.

REFERENCES

[1] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.

[2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[3] O. Ledoit and M. Wolf, "Honey, I shrunk the sample covariance matrix," *J. Portfolio Manage.*, vol. 30, no. 4, pp. 110–119, 2004.

[4] K. Sheppard, "arch: Autoregressive conditional heteroskedasticity models in Python," https://arch.readthedocs.io/, accessed Dec. 2025.

[5] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," in *Proc. 9th Python Sci. Conf.*, 2010.