ARMA–GARCH–MNTS CVaR Portfolio Optimization: An Out-of-Sample 2025 Backtest on U.S. Equities

Abstract

This project develops an equity portfolio construction pipeline that combines ARMA–GARCH filtering, a Multivariate Normal Tempered Stable (MNTS) model for multi-asset returns, and CVaR-based optimization of 21-day loss. Using 30 large-cap U.S. equities from roughly 2019–2024 as the training sample, we build three portfolios: a historical mean–variance max-Sharpe portfolio (MS_historical) and two MNTS-based CVaR max-Sharpe portfolios, one with ARMA-filtered residuals (MS_CVaR_ARMA) and one without ARMA dynamics (MS_CVaR_noARMA). We then evaluate the out-of-sample performance of these static buy-and-hold portfolios, holding the 31-Dec-2024 weights fixed across ten non-overlapping 21-day windows in 2025. The objective is to test the robustness of the initial allocation, not a frequently rebalanced trading strategy. Out-of-sample, the CVaR portfolios deliver slightly higher average 21-day returns and clearly lower realized volatility than the historical max-Sharpe portfolio, while generating zero 95% VaR breaches. MNTS-based predictions are risk-conservative: simulated volatility is higher than realized volatility, but the realized 21-day returns fall near the center of the simulated distribution, suggesting reasonable shape calibration. Differences between the ARMA and no-ARMA CVaR portfolios are modest: their predicted risk metrics, realized performance, percentile distributions, and VaR behavior are all very similar. Overall, the results support MNTS-based CVaR optimization as a viable, more downside-aware alternative to standard mean–variance optimization for this universe and horizon.

1. Introduction

This project builds and tests a portfolio construction pipeline that combines:

- ARMA–GARCH models for conditional volatility,
- a Multivariate Normal Tempered Stable (MNTS) distribution for joint returns,
- and CVaR-based optimization for downside-risk control.

The goal is to compare a traditional historical mean–variance max-Sharpe portfolio with two CVaR-optimized portfolios built on MNTS simulations:

1. MS_historical – max-Sharpe portfolio using historical mean and covariance.

2. MS_CVaR_ARMA – max-Sharpe on a CVaR frontier using MNTS scenarios with ARMA-filtered residuals.

3. MS_CVaR_noARMA – Same as (2) but using MNTS on raw returns (no ARMA dynamics).

The main questions are:

- How different are the predicted risk/return profiles from the MNTS model vs. what actually happens out-of-sample?

- Do the CVaR portfolios improve the risk-return trade-off relative to the historical max-Sharpe portfolio?

- Is the MNTS model reasonably calibrated in terms of volatility, percentiles, and VaR breaches?

2. Data and universe

- Universe: a set of 30 large-cap US equities (Dow-style names such as AAPL, AMGN, etc.).

- Prices: daily adjusted close prices obtained via tq_get (tidyquant) from a public data provider.

- Returns: log returns computed from adjusted prices.

The pipeline uses a training window from approximately 2019 through the end of 2024 (as configured in config.yml, e.g., start_date to end_date = 2024-12-31).

The out-of-sample period is the calendar year 2025, split into 10 non-overlapping 21-trading-day windows. Each window is treated as a 21-day horizon for realized performance.

3. Modeling and simulation

3.1 ARMA–GARCH filtering

For each asset:

1. Fit an ARMA(p, q) model to daily log returns.
2. Fit a GARCH-type volatility model to the residuals.
3. Standardize residuals to approximate i.i.d. noise.

This step is used in the ARMA version of the MNTS model. In the no-ARMA version, MNTS is fitted directly to the raw returns.

3.2 MNTS fit and scenario generation

Using the standardized residuals (or raw returns):

- Fit a Multivariate Normal Tempered Stable (MNTS) distribution to the cross-section of returns.
- Using the fitted parameters, simulate a large number of 21-day scenarios:
    - Example: 20,000 21-day return vectors across the 30 assets.
    - Stored as sims_mnts_21d.rds (ARMA) and sims_mnts_21d_noarma.rds (no-ARMA).

From these simulations, we obtain:

- Simulated 21-day means and covariances,
- Full loss distributions per portfolio,
- Empirical 95% VaR and CVaR for any chosen weight vector.

4. Portfolio construction

4.1 Historical max-Sharpe (MS_historical)

Figure 1 illustrates the historical mean–variance efficient frontier and the selected max-Sharpe portfolio used as the baseline (MS_historical).

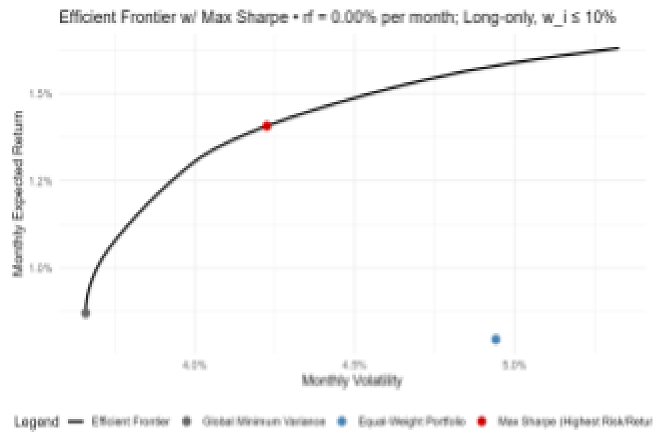

Figure 1: Mean–variance efficient frontier and selected historical max-Sharpe portfolio (2019–2024 training sample).

Using the in-sample period up to 2024-12-31:

- Compute historical mean vector and covariance matrix of returns.
- Solve a long-only max-Sharpe problem with a per-asset weight cap (e.g., 10%).
- The resulting weights are stored in outputs/tables/weights_max_sharpe.csv.

This is the baseline portfolio.

4.2 CVaR frontier portfolios

Figure 2 shows the MNTS-based CVaR frontiers for the ARMA and no-ARMA specifications, together with the corresponding max-Sharpe portfolios (MS_CVaR_ARMA and MS_CVaR_noARMA).

Using the MNTS scenarios, we construct a CVaR frontier in weight space:

- Decision variables: asset weight vector w (long-only, sum to 1, with caps).
- Objective: minimize 95% CVaR of 21-day loss, subject to a target expected 21-day mean return.
- This is implemented via the Rockafellar–Uryasev formulation in CVXR:

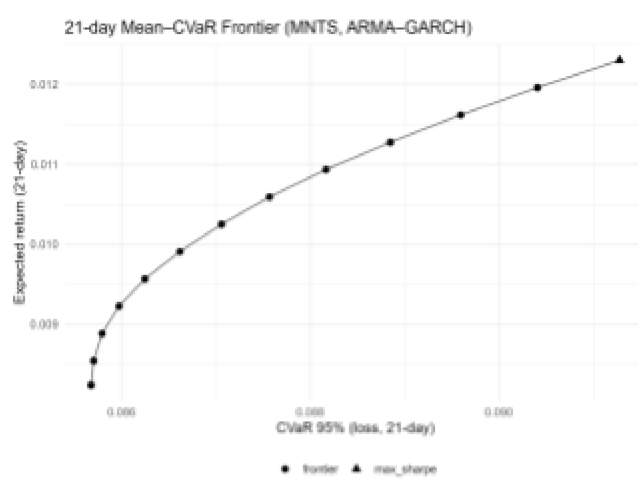

Figure 2: MNTS-based 21-day CVaR frontier in mean–standard-deviation

- Introduce VaR threshold t and slack variables u_s for each scenario s.
- $Minimize\ t + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} u_s$

  subject to: $u_s \geq 0\ \ and\ \ u_s \geq L_s(w) - t$

By sweeping a grid of target means, we obtain a frontier of CVaR-efficient portfolios. For each point on the frontier, we compute:

- Expected 21-day mean return,
- Standard deviation,
- 95% VaR and CVaR,
- Sharpe ratio (mean / SD),
- Full set of asset weights.

We then select the max-Sharpe portfolio on this CVaR frontier, separately for:

- MNTS with ARMA (saved in frontier_cvar_21d.csv),
- MNTS without ARMA (saved in frontier_cvar_21d_noarma.csv).

The corresponding weight vectors define the portfolios:

- MS_CVaR_ARMA
- MS_CVaR_noARMA

5. Out-of-sample evaluation methodology (static 2025 test)

In this report, the 2025 analysis is intentionally static.

We fix the three weight vectors estimated on the in-sample period up to 2024-12-31 (MS_historical, MS_CVaR_ARMA, MS_CVaR_noARMA) and then evaluate how those unchanged portfolios perform over 2025.

This should be interpreted as an out-of-sample robustness check of the initial allocation, not as a fully dynamic trading strategy with periodic re-optimization.

1. Fix the three portfolio weight vectors at their 2024-12-31 values and carry them forward unchanged through 2025.

2. For 2025, fetch daily adjusted prices for all tickers and compute log returns.
3. Construct 10 non-overlapping 21-trading-day windows (roughly Jan–Nov 2025).

4. For each window and each portfolio:

    ◦ Compute the daily portfolio log return as $r_t^{(p)} = w^\top r_t$ where r_t is the vector of asset log returns.

    ◦ Aggregate to 21-day log return by summing over the window.

Across the 10 windows, we compute:

- Realized mean of 21-day log returns,

- Realized SD of those 10 window returns,

- For each realized 21-day return, its percentile within the simulated MNTS distribution,

- Whether the realized 21-day loss exceeded the simulated 95% VaR.

The summary statistics are stored in:

- outputs/tables/backtest_2025_summary.csv – aggregate stats per portfolio.

- outputs/tables/backtest_2025_windows_detail.csv – per-window returns, percentiles, and VaR breaches.

6. Results

6.1 Aggregate predicted vs realized metrics

Table 1 reports, for each portfolio:

Metric	MS_historical	MS_CVaR_ARMA	MS_CVaR_noARMA
Predicted mean (21-day)	0.01375247	0.01230011	0.01226047
Predicted SD (21-day)	0.05348370	0.04863403	0.04858372
Predicted 95% VaR loss (21-day)	0.07312154	0.06692282	0.06695838
Predicted 95% CVaR loss (21-day)	0.10131631	0.09128042	0.09119113
Realized mean (21-day)	0.007185979	0.010505739	0.010504103
Realized SD (21-day)	0.04011181	0.02665810	0.02683079
Avg percentile in MNTS sims	0.464755	0.497040	0.497475
VaR95 breach rate	0	0	0

- Predicted 21-day mean and SD from MNTS,
- Predicted 95% VaR and CVaR of 21-day loss,
- Realized mean and SD across the 10 windows,
- Average percentile of realized returns in the simulated distribution,
- Fraction of windows with a VaR breach.

Key patterns:

- All three portfolios have similar predicted 21-day mean returns from the MNTS sims.
- The CVaR portfolios exhibit slightly higher realized mean 21-day return than the historical portfolio over 2025.
- Realized volatility is lower than predicted for all three portfolios, with the largest reduction for the CVaR portfolios.
- The empirical VaR breach rate is 0% for all portfolios in 2025 (no window has a 21-day loss exceeding the simulated 95% VaR).

6.2 Realized Sharpe and drawdowns

Portfolio	Annualized Sharpe (2025)	Max drawdown in 2025
MS_historical	0.62	-8.9%
MS_CVaR_ARMA	1.52	-5.0%
MS_CVaR_noARMA	1.36	-5.1%

Table 2 summarizes realized risk-adjusted performance in 2025.

- The two CVaR portfolios clearly dominate the historical max-Sharpe benchmark on a Sharpe basis: realized annualized Sharpe is about 0.62 for MS_historical vs 1.52 (MS_CVaR_ARMA) and 1.36 (MS_CVaR_noARMA).
- Maximum drawdowns are also smaller for the CVaR portfolios: about -8.9% for MS_historical vs -5.0% and -5.1% for the CVaR portfolios.
- Combining higher Sharpe and smaller drawdowns, the MNTS-based CVaR optimization appears to deliver a more stable risk-return profile over this 2025 out-of-sample period, even with static weights.

These results are consistent with the rest of the figures: the CVaR portfolios exhibit lower realized volatility (Section 6.3) and smoother paths across the ten windows (Section 6.4).

6.3 Realized 21-day returns across windows

Figure 3 tracks the 21-day log returns for each portfolio across the 10 windows in 2025.

Observations:

- All portfolios respond to the same market cycles: a sharp drawdown early in the year, followed by recoveries and another mid-year dip.
- The historical max-Sharpe portfolio tends to show the largest swings, with the deepest drawdown and the largest positive spikes.
- The CVaR portfolios are smoother:
    - Maximum 2025 drawdown is about -8.9% for the historical portfolio versus -5.0% and -5.1% for the two CVaR portfolios, so downside excursions are roughly 3–4 percentage points smaller under CVaR optimization.
    - Benchmark-level upside is still captured; in several windows, the CVaR portfolios slightly outperform the historical portfolio.

Figure 3: Realized 21-day log returns for the three portfolios across ten non-overlapping windows in 2025.

Overall, the realized paths suggest that CVaR optimization successfully trades some extreme exposures for more stable outcomes, without sacrificing average 21-day return.

6.4 Predicted vs realized volatility

Figure 4 compares predicted vs realized 21-day standard deviation for each portfolio.

- MNTS simulations overestimate volatility:
    - Predicted SDs are clearly higher than realized SDs.
- For the CVaR portfolios, the gap between predicted and realized volatility is largest, meaning the model was particularly

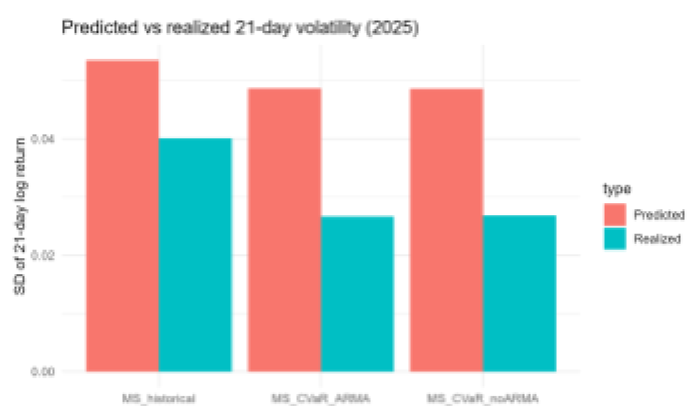

Figure 4: Predicted vs. realized 21-day volatility in 2025 for the historical max-Sharpe and two CVaR portfolios.

conservative about their downside risk.

- The ranking is consistent:
    - Historical max-Sharpe remains the most volatile,
    - CVaR portfolios deliver lower realized volatility.

Interpretation: The MNTS model is risk-conservative in 2025; it errs on the side of overstating risk, particularly for risk-managed portfolios.

6.5 Distribution of realized percentiles

Figure 5 shows the distribution of realized percentiles of each 21-day return inside the simulated MNTS distribution.

- If the model is well-calibrated, these percentiles should scatter around 0.5.
- For all three portfolios, the median percentile is close to 0.5, and the interquartile ranges are not heavily skewed toward the tails.
- No indication of realized returns concentrating in either tail of the simulated distribution.

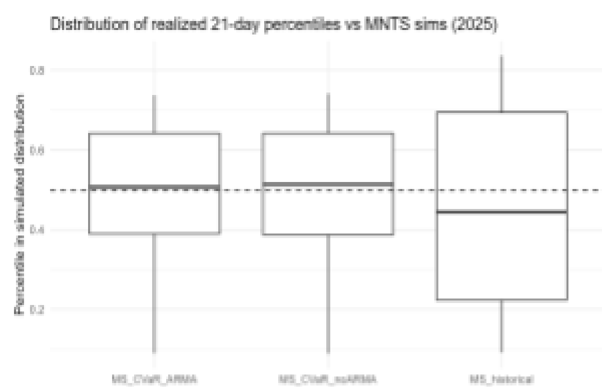

Figure 5: Distribution of realized 21-day return percentiles within the MNTS simulated distributions in 2025.

Given only 10 windows, this is a small sample, but it suggests that the MNTS model's shape is reasonably consistent with realized outcomes.

6.6 VaR breach rate

Figure 6 plots the empirical breach rate of 95% VaR for each portfolio, alongside the theoretical 5% benchmark.

- The observed breach rate is 0% for all three portfolios.

Figure 6: Empirical 95% VaR breach rate in 2025 for each portfolio, compared with the theoretical 5% benchmark.

- With only 10 windows, a 0% breach rate is statistically plausible given such a small sample, but combined with the volatility overprediction, it reinforces the idea that the MNTS-based VaR is conservative for this period.

From a risk-management standpoint, this is preferable to the opposite error (underestimating tail risk).

7. Comparison of ARMA vs no-ARMA CVaR portfolios

The two CVaR portfolios behave very similarly:

- Their predicted mean, SD, VaR, and CVaR are close.
- Their realized returns and volatilities across the 10 windows are almost indistinguishable in the plots.
- Percentile distributions and VaR breach behavior are also very similar.

This suggests that, at least for this universe and horizon, adding ARMA dynamics into the MNTS modeling provides only incremental differences relative to the simpler no-ARMA specification. Both versions deliver comparable performance and risk control.

8. Limitations and possible extensions

Several limitations are worth noting:

1. Sample size of windows
   Only 10 non-overlapping 21-day windows are available in 2025. Though sufficient for a basic sanity check, not enough to precisely estimate breach frequencies or calibration errors.
2. Single market regime
   The backtest covers one year. If 2025 happens to be relatively calm, conservative risk estimates will naturally look very safe. A longer out-of-sample period with multiple

regimes would give stronger evidence.

3. Transaction costs and turnover

    Static, non-rebalanced portfolios. The 2025 analysis keeps the Dec-2024 weights fixed and evaluates their performance over ten non-overlapping 21-day windows. This is a robustness test of the initial allocation, not a full rebalancing strategy. A more realistic implementation would re-run the ARMA–GARCH–MNTS pipeline every 21 days (or monthly) and compare the dynamic CVaR strategy to a dynamically rebalanced historical benchmark.

4. Model risk

    MNTS is flexible but still a parametric model. Mis-specification, fitting error, and numerical issues can all affect the simulated distribution. Non-parametric or semi-parametric alternatives (e.g., filtered historical simulation) could be used as comparisons.

Potential extensions:

- Rolling or expanding-window re-estimation of the MNTS and GARCH parameters.
- Testing other risk measures (e.g., Expected Shortfall at different confidence levels).
- Comparing MNTS-based portfolios to simpler Gaussian or t-copula models on the same universe.
- Stress testing with scenario shocks applied to the MNTS parameters.

9. Conclusion

This project implemented an end-to-end pipeline that:

1. Ingests daily equity data and computes log returns,
2. Fits ARMA–GARCH and MNTS models to generate multi-asset 21-day scenarios,
3. Builds a historical max-Sharpe portfolio and MNTS-based CVaR frontier portfolios, and
4. Backtests these strategies on non-overlapping 21-day windows in 2025.

The main empirical findings are:

- The CVaR portfolios deliver slightly higher average 21-day returns with lower realized volatility than the historical max-Sharpe portfolio.

- MNTS-based predictions are conservative: they over-predict volatility and produce zero 95% VaR breaches during 2025.

- The simulated distribution appears reasonably calibrated, with realized returns landing near the middle of the MNTS distribution rather than in the tails.

- The difference between the ARMA and no-ARMA MNTS specifications is modest; both CVaR portfolios behave similarly in practice.

Overall, the results support the use of MNTS-based CVaR optimization as a viable alternative to standard mean–variance optimization for this universe and horizon, particularly from a risk-management perspective.