

Tail-risk-constrained portfolio optimization with cluster pre-selection and machine learning: risk-adjusted evidence from the S&P 500

Journal of Capital
Markets Studies

David Toscano and Juan C. Roca

Department of Financial Economics and Accounting, Faculty of Business Sciences
and Tourism, University of Huelva, Huelva, Spain

Received 22 June 2026
Revised 26 September 2026
Accepted 26 September 2026

Abstract

Purpose – Standard Gaussian risk frameworks underestimate extreme joint losses in large equity universes, exposing investors to unbudgeted drawdowns. This article develops and tests a portfolio-construction pipeline that disciplines tail risk while exploiting forward-looking return signals.

Design/methodology/approach – The four-stage pipeline integrates a rolling tail-index (Hill) filter, cluster pre-selection, a Conditional Value-at-Risk (CVaR) risk-parity optimizer with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and extreme value theory (EVT) marginals and a Student-t copula, and an eight-classifier machine-learning ensemble that reallocates weights within the risk budget. It is tested on S&P 500 constituents across 75 monthly walk-forward windows (January 2020–January 2026), net of transaction costs.

Findings – Among twelve strategies, the pipeline delivers the highest Sharpe ratio (1.106) and monthly win rate (76.0%), surpassing risk parity, Hierarchical Risk Parity and the Gaussian CVaR specification it generalizes; block-bootstrap inference confirms its advantage over passive exposure (resampled mean Sharpe 1.316). Risk parity accumulates greater terminal wealth, a trade-off, since tail constraints curb high-momentum exposure in bull markets.

Research limitations/implications – The analysis is limited to S&P 500 constituents and a January 2020–January 2026 evaluation period. Results may differ in other asset classes, international markets or under alternative transaction-cost assumptions. Future research could examine adaptive clustering schemes, alternative dependence structures and real-time implementation constraints.

Practical implications – For investors prioritizing risk-adjusted efficiency over cumulative wealth, the framework is a transparent capital-allocation tool whose tail discipline holds at roughly 97 basis points annually.

Originality/value – To the authors' knowledge, this is the first to combine tail-index screening, clustering, GARCH-EVT-copula CVaR optimization and a machine-learning overlay within one framework, jointly addressing the dimensionality, distributional misspecification and information-neglect problems that defeat these tools in isolation.

Keywords Tail risk, Portfolio optimization, Extreme value theory, Conditional value-at-risk, Machine learning, S&P 500

Paper type Research article

1. Introduction

In March 2020, the S&P 500 lost more than 30% of its value in twenty-three trading days. Risk management systems at many institutional desks continued to display green dashboards throughout the decline. While these models were functioning correctly, they were measuring the wrong risk. According to a standard Gaussian framework, a loss of that magnitude was roughly as likely as a coin landing heads thirty times in a row. Nobody in a trading room

JEL Classification — G11, C58, C45, C38

© David Toscano and Juan C. Roca. Published in *Journal of Capital Markets Studies*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licences/by/4.0/>

Funding: This research was supported by the Research Group SEJ-640 “Empirical Economics and Finance” (ECOFINE), registered in the Andalusian Scientific Information System (SICA).



Journal of Capital Markets Studies
Emerald Publishing Limited
2514-4774
DOI 10.1108/JCMS-06-2026-0114

believed such an outcome, yet those were the numbers driving their allocations. We take this gap between what practitioners sense and what models report as our starting point.

The failure of Gaussian risk models under market stress is not a new finding. However, what makes this issue practically urgent is the context in which it recurs: an institutional investor attempting to manage a diversified equity portfolio at scale, selecting a subset of S&P 500 companies that must simultaneously limit extreme losses and capture the return predictability documented in the literature. This task involves three distinct challenges that standard frameworks rarely solve simultaneously. The first is a dimensionality problem. Classical mean-variance optimization becomes numerically unstable in large asset universes. [DeMiguel et al. \(2009\)](#) show that the estimation window required to outperform a naive equal-weight benchmark grows to thousands of months as the number of assets increases: with so many parameters to estimate, the sample covariance matrix accumulates more estimation error than the optimization can offset, so any theoretical gain is lost well before it can be realized. This problem persists across the full breadth of tradable firm characteristics ([DeMiguel et al., 2020](#)). Cluster-based pre-selection directly addresses this failure by reducing the investable universe to a well-conditioned subset, thereby stabilizing out-of-sample performance ([Cirulli et al., 2024](#); [Lorenzo and Arroyo, 2023](#)).

However, dimensionality reduction alone does not address the other two challenges. The second is a distributional problem. Even a well-estimated covariance matrix fails when we assume normality. Gaussian Conditional Value-at-Risk (CVaR) systematically underestimates the probability of extreme joint losses ([Artzner et al., 1999](#); [Rockafellar and Uryasev, 2000](#)). Non-Gaussian estimation improves drawdown protection under market stress ([McNeil and Frey, 2000](#); [Paolella et al., 2025](#)), yet most frameworks continue to rely on Gaussian assumptions out of computational convenience. The third is a temporal problem. Standard risk-parity frameworks ([Maillard et al., 2010](#)) allocate based on historical distributions and discard the forward-looking information embedded in equity return cross-sections. Machine learning methods can extract return-predictive signals from large panels of firm characteristics ([Gu et al., 2020](#); [Kelly et al., 2024](#); [Sheng et al., 2025](#)), but they are rarely integrated with a coherent tail-risk framework.

Each of these problems has attracted partial solutions, but no study resolves all three within a single, internally consistent framework. Prior work applying machine learning to portfolio selection rarely models tail dependence ([Bradrania et al., 2022](#); [Bryzgalova et al., 2025](#); [Rigamonti, 2024](#)), and research combining copula, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and CVaR typically operates on cryptocurrency or small international samples rather than S&P 500 constituents ([Mba et al., 2018](#); [Mba and Mwambi, 2020](#); [Simo-Kengne et al., 2018](#)). To the best of our knowledge, this is the first study to integrate tail-index filtering, cluster-based pre-selection, a GARCH-EVT copula CVaR risk-parity optimizer and an eight-classifier machine-learning ensemble in one pipeline, tested on S&P 500 constituents over a full market cycle. We make three contributions. First, we show that distributional accuracy in the tail is what preserves capital during stress, while the machine-learning overlay adds return without enlarging drawdown. Second, we demonstrate that cluster-based pre-selection is a structural precondition rather than a computational convenience: removing it collapses risk-adjusted performance below the passive index. Third, we provide block-bootstrap evidence that the pipeline's risk-adjusted advantage over passive exposure is statistically robust, not an artefact of a single sample path.

We propose and test such a framework using S&P 500 companies across 75 monthly walk-forward windows from January 2020 to January 2026. This period covers the COVID-19 collapse, the 2022 rate-driven bear market and two distinct bull phases. We choose the S&P 500 because it is the primary institutional equity benchmark globally, large enough to expose the limits of Gaussian assumptions and liquid enough that the tail events we document are structurally representative rather than episodic. We posit that the three failure modes identified above are hierarchical: attempting to model temporal return predictability is futile if the underlying risk model is misspecified and extreme value theory (EVT)-based risk estimation is

unreliable if the input universe is corrupted by dimensional noise. Each analytical component is therefore a theoretical precondition for the validity of the next.

We structure our approach around four sequential stages. In Stage 0, we apply a rolling [Hill \(1975\)](#) estimator to each constituent's return series and retain only those stocks with a stable tail index, approximately 220 per window. In Stage 1, the surviving universe is partitioned via the Clustering Large Applications (CLARA) algorithm ([Kaufman and Rousseeuw, 1990](#)), resulting in a low-dimensional, well-conditioned investable subset. In Stage 2, we construct a CVaR-risk-parity portfolio using GARCH(1,1)–EVT marginals and a Student-*t* copula to capture joint tail dependence. In Stage 3, we reallocate weights using an ensemble of eight machine-learning classifiers, operating strictly within the risk budget established in Stage 2.

Our full strategy attains a Sharpe ratio of 1.106 and a monthly win rate of 76.0%, both the highest in the sample, with a maximum drawdown of -27.4% and annualized volatility of 21.8%. Risk parity (Sharpe 1.089) and Hierarchical Risk Parity (HRP, Sharpe 1.064) are the strongest conventional comparators but require higher annualized volatility (24.5 and 24.0% respectively) and achieve substantially lower win rates (69.3 and 68.0%). Passive S&P 500 exposure delivers a Sharpe of 0.776, providing the natural lower bound for any active strategy. Relative to passive S&P 500 exposure, the pipeline adds 9.8% points of annualized return (24.1 versus 14.3%) while deepening the maximum drawdown by only 4.4 points, and block-bootstrap resampling confirms its advantage over passive exposure (resampled mean Sharpe 1.316). The pipeline's edge is therefore one of risk-adjusted efficiency and consistency rather than raw accumulation: it posts the highest Sharpe ratio and the highest monthly win rate of any strategy we test.

We confirm these results through a set of controlled comparisons that withhold one component at a time from the integrated pipeline, leaving the shared eligible universe in place. Replacing the full GARCH-EVT-copula risk model with a Gaussian CVaR specification raises the point Sharpe slightly, from 0.988 to 1.080, but widens the maximum drawdown from -27.4% to -28.6% and lowers the monthly win rate from 74.7% to 66.7%, which indicates that distributional accuracy matters more for loss control than for average return. Adding the machine-learning weight adjustment then raises the Sharpe from 0.988 to 1.106 without widening tail risk, confirming that the predictive layer contributes return without disrupting the risk structure achieved in the previous stage. The most consequential result concerns the cluster-based pre-selection step, which reduces the 451-asset universe to a well-conditioned investable subset before optimization. Removing it and selecting assets directly by Sharpe ranking collapses the Sharpe ratio to 0.649, below the passive index and close to a naive equal-weight benchmark. Pre-selection is not a computational convenience but a structural precondition for the gains documented in later stages.

We organize the remainder of this article as follows. In [Section 2](#), we review the related literature on portfolio risk measures and CVaR-based risk parity, tail-risk estimation with GARCH dynamics and EVT, machine learning for equity return prediction and asset pre-selection in large equity universes. In [Section 3](#), we describe our four-stage methodology. In [Section 4](#), we report our empirical results, including the contribution of each pipeline component and bootstrap robustness checks. In [Section 5](#), we conclude.

2. Literature review

2.1 Portfolio risk measures: from mean-variance to CVaR-based risk parity

The mean-variance framework of [Markowitz \(1952\)](#) remains the canonical reference for portfolio optimization. Its Gaussian distributional assumptions, however, cause it to systematically underestimate tail risk during market dislocations. [Artzner et al. \(1999\)](#) demonstrate that Value-at-Risk fails to satisfy subadditivity and propose a coherent risk framework whose requirements CVaR satisfies; [Rockafellar and Uryasev \(2000\)](#) establish that CVaR can be minimized via linear programming, making it computationally tractable for large-scale portfolio construction.

The risk parity framework of [Maillard et al. \(2010\)](#) sidesteps return forecasting by equalizing risk contributions across assets, thereby reducing the estimation error that [DeMiguel et al. \(2009\)](#) identify as the primary reason optimized strategies fail to outperform the naïve $1/N$ rule out of sample. The natural extension replaces variance with CVaR as the equalized metric. [Sheng et al. \(2025\)](#) propose a CVaR-based risk parity model augmented with a machine learning layer and report a Sharpe ratio approximately three times that of equal-weight benchmarks. [Kaur et al. \(2026\)](#) confirm that ML-assisted CVaR optimization outperforms variance-based benchmarks across bull, bear and volatile regimes. A limitation common to both studies is retained Gaussian distributional assumptions; [Sheng et al. \(2025\)](#) explicitly acknowledge that their model may underestimate tail risk under actual market conditions. [Section 3](#) of the present article addresses this limitation directly.

2.2 Tail risk estimation: GARCH dynamics, extreme value theory and tail dependence

Two well-documented regularities in financial return distributions motivate the estimation framework adopted here. Volatility clustering, the tendency of large movements to follow large movements, was characterized by [Engle \(1982\)](#) and generalized by [Bollerslev \(1986\)](#). Fat tails and excess kurtosis have been documented across markets and sample periods since [Fama \(1965\)](#) and formalized by [Embrechts et al. \(1997\)](#). GARCH models with conditionally normal errors capture time-varying volatility but impose normality on standardized residuals; [McNeil and Frey \(2000\)](#) show that this assumption does not hold and that the most dangerous losses occur precisely where the Gaussian assumption is most wrong.

[McNeil and Frey \(2000\)](#) resolve this with a two-step conditional EVT approach: GARCH filtering is applied first; the Peaks-over-Threshold method, which fits a Generalized Pareto Distribution to exceedances following [Pickands \(1975\)](#) and [Balkema and de Haan \(1974\)](#), is then applied to the approximately i.i.d. standardized residuals. Rigorous backtesting confirms that conditional EVT consistently outperforms both GARCH-with-normal-errors and unconditional EVT for expected shortfall estimation.

At the portfolio level, [Longin and Solnik \(2001\)](#) document that cross-market correlations increase sharply in bear markets and that co-movements in the extreme left tail far exceed Gaussian predictions. [Embrechts et al. \(2002\)](#) show that the Student- t copula captures positive symmetric tail dependence that Gaussian copulas cannot represent. Recent multi-asset evidence confirms this superiority: [Sahamkhadam and Stephan \(2023\)](#) find that mixed vine copula specifications consistently reduce portfolio tail risk across a 21-year international sample; [Steinborn and Liebscher \(2025\)](#) validate the framework in an equal risk contribution context spanning equities, bonds, hedge funds and catastrophe bonds. [Lee \(2025\)](#) further shows that retaining a GARCH layer is essential even when deep learning is available, as the GARCH component anchors tail probability forecasts to current volatility conditions.

2.3 Machine learning for equity return prediction and dynamic portfolio adjustment

Machine learning has generated documented economic gains over linear factor models for equity return prediction. [Gu et al. \(2020\)](#), covering the U.S. common stock universe from 1957 to 2016, find that tree-based methods and neural networks consistently outperform penalized linear models by capturing nonlinear interactions among momentum, liquidity and volatility signals; [Leippold et al. \(2022\)](#) confirm that these gains generalize beyond the US context.

Within the binary return classification framework this article adopts, [Krauss et al. \(2017\)](#) show that ensemble combinations of deep neural networks, gradient-boosted trees and random forests outperform individual classifiers on S&P 500 constituents. [Fischer and Krauss \(2018\)](#) document that standalone ML alpha has declined since approximately 2010, a finding [Wolff and Echterling \(2024\)](#) confirm on S&P 500 and STOXX Europe 600 data, while showing that regularized logistic regression performs statistically indistinguishably from long short-term memory (LSTM) networks and gradient-boosted trees, a result that motivates ensemble voting across architectures rather than selection of a single dominant model. [Htun et al. \(2024\)](#)

confirm that binary classification signals persist in S&P 500 data from 2017 to 2022 with LSTM outperforming alternative classifiers at a ten-day horizon.

The erosion of standalone ML alpha argues for deploying machine learning within a structurally sound optimization framework rather than as a pure signal-generation layer. [Sarkar et al. \(2025\)](#) show that embedding ML prediction within a portfolio objective function outperforms sequential two-stage approaches. [Sheng et al. \(2025\)](#) formalize this architecture using an eight-classifier ensemble that generates majority-vote predictions to adjust CVaR-RP weights dynamically, drawing on [Gu et al.'s \(2020\)](#) finding that no single classifier dominates across all market regimes. The present study adopts this architecture and extends it to a GARCH-EVT tail-estimated optimizer, ensuring that ML weight adjustments provide incremental forward-looking information on top of an already sound risk allocation.

2.4 Asset pre-selection and dimensionality reduction in large equity universes

Mean-variance optimization over N assets requires estimating $N(N + 1)/2$ covariance parameters. [DeMiguel et al. \(2009\)](#) demonstrate that estimation error systematically overwhelms optimization gains at realistic sample sizes: a 25-asset portfolio requires approximately 3,000 months of data to outperform $1/N$, rising to 6,000 for 50 assets. For an S&P 500-scale universe, pre-selection is therefore a statistical necessity rather than a computational convenience.

Cluster-based pre-selection provides a principled response. Applying CLARA clustering to the bivariate mean-return/volatility space, [Lorenzo and Arroyo \(2023\)](#) identify investable subsets of cryptocurrencies whose risk-adjusted performance substantially exceeds full-universe optimization. [Paolella et al. \(2025\)](#) show that non-Gaussian risk parity generates materially lower turnover during stress periods, confirming that selecting assets whose tail properties match the estimation model improves both performance and stability.

2.5 Research gap

We evaluate the two proposed strategies against 10 benchmarks. Because the contribution of this article is the integration of four components that prior work has used only in isolation, the benchmarks serve two purposes. First, they establish that the integrated pipeline is worth building: the S&P 500 ETF, two naive diversification rules and the two strategies of [Lorenzo and Arroyo \(2023\)](#) test whether the pipeline improves on passive investing, on simple diversification and on the prior literature, respectively. Second, they confirm that the integration is not gratuitous: by withholding one component at a time, we verify that each piece earns its place in the whole. The tail-index filter of Stage 0 is common to every strategy below, since it defines the eligible universe from which all of them draw, so it acts as the shared foundation of the comparison rather than as a treatment to be switched on or off; the remaining three components are isolated by dedicated comparisons. [Table 4](#) records, for every strategy, which pipeline stages it activates and which asset universe it uses.

3. Methodology

3.1 Data and sample

We use daily adjusted closing prices for 451 S&P 500 constituents from Yahoo Finance, covering January 2, 2018 through March 30, 2026, which yields approximately 921,922 ticker-day observations. Returns are continuously compounded log-returns, $r_{it} = \log(P_{it}/P_{i,t-1})$, where P_{it} denotes the split- and dividend-adjusted closing price of asset i on day t .

We split the sample into a burn-in phase (January 2018 to December 2019) and an out-of-sample evaluation phase of 75 monthly walk-forward windows spanning January 2020 through January 2026. At each investment date t , we estimate all model parameters on the 730-calendar-day window ending on t and we hold the portfolio for 30 calendar days before the next rebalancing. No future information enters any estimation or classification step. We deduct

transaction costs at 5 basis points per unit of one-way portfolio turnover, $T_t = \frac{1}{2} \sum_i |w_{it} - w_{i,t-1}^+|$, following [DeMiguel et al. \(2009\)](#).

[Table 1](#) reports cross-sectional summary statistics for the 451 constituents over the full sample. All 451 assets reject the Jarque-Bera null of normality at the 5% level, confirming the fat-tailed, negatively skewed return distributions that motivate the GARCH-EVT-copula specification in Stage 2.

3.2 Framework overview

The proposed framework applies four sequential stages at each monthly rebalancing date. Stage 0 screens the 451-asset universe for tail-eligible assets using a rolling Hill estimator. Stage 1 clusters the eligible assets in mean-return/volatility space via the CLARA algorithm and retains the highest-Sharpe assets from each valid cluster. Stage 2 constructs a CVaR-risk-parity portfolio over the selected assets, combining GARCH(1,1)-EVT marginals with a Student- t copula for joint tail dependence. Stage 3 adjusts Stage 2 weights by excluding assets whose 30-day return direction is predicted negative by an eight-classifier ensemble. [Sections 3.3–3.6](#) detail each stage in turn.

3.3 Stage 0: tail-index filter

The Stage 2 optimizer requires assets with finite variance and nontrivial variability: infinite-variance tails yield ill-conditioned GARCH estimates, and near-static assets attract disproportionately large risk-parity weights. Stage 0 therefore applies three sequential filters at each 730-day estimation window.

3.3.1 Hill estimator. We estimate the tail index for each asset i from the $k = 50$ upper order statistics of the absolute return series, following [Hill \(1975\)](#). Let $X_1 \geq X_2 \geq \dots \geq X_n$ denote the ordered tail observations. The estimator is:

$$\hat{\xi}_i = \frac{1}{k} \sum_{j=1}^k \log(X_j) - \log(X_{k+1}), \quad \hat{\alpha}_i = 1 / \hat{\xi}_i. \tag{1}$$

We compute $\hat{\alpha}_i^+$ and $\hat{\alpha}_i^-$ separately for positive and negative return tails, and define $\hat{\alpha}_i^{\min} = \min(\hat{\alpha}_i^+, \hat{\alpha}_i^-)$. A minimum of 60 valid tail observations is required.

3.3.2 Eligibility filters. Asset i enters the eligible universe $\mathcal{E}_t$ for window t only if it meets three conditions at once. Its data coverage must be adequate, so at least 90% of closing prices within the estimation window are nonmissing (F1). Its tail index $\hat{\alpha}_i^{\min}$ must weakly exceed both the cross-sectional median across all assets and an absolute floor of 1.5, following [Lorenzo and Arroyo \(2023\)](#) (F2). Its annualized return volatility must also exceed 10%, which rules out near-static assets (F3).

Table 1. Cross-sectional summary statistics for 451 S&P 500 constituents

Statistic	Mean	Std	P25	Median	P75	Min	Max
Annualized return	6.0%	10.0%	0.3%	6.4%	11.4%	−27.8%	48.8%
Annualized volatility	35.3%	10.6%	27.7%	32.4%	40.6%	19.4%	84.4%
Skewness	−0.63	1.06	−0.89	−0.43	−0.14	−13.53	1.64
Excess kurtosis	16.54	24.17	7.68	11.83	17.40	2.56	417.14
JB test (% rejecting normality at 5%)	100.0%	—	—	—	—	—	—

Note(s): Statistics computed over the full sample (January 2, 2018–March 30, 2026). Returns are continuously compounded daily log-returns; annualized figures use 252 trading days. Excess kurtosis = kurtosis −3. JB: percentage of assets rejecting the Jarque-Bera null of normality at the 5% significance level

Filter F2 is adaptive: its threshold tightens automatically during market stress, when cross-sectional tails are heavier, and relaxes in calm periods. Over the 75 evaluation windows, it retains on average 222.1 assets (range: 218–226). Table 2 reports the asset funnel across the successive Stage 0 filters. Table 3 summarizes the characteristics of the retained assets, confirming that the eligible universe consists of assets with well-behaved tail distributions.

3.4 Stage 1: CLARA clustering

Stage 1 reduces the eligible universe to a compact investable set, following Lorenzo and Arroyo (2023), by retaining the cluster whose Sharpe prototype offers the most attractive risk-adjusted profile and eliminating assets with weak risk-return characteristics before the Stage 2 optimization.

3.4.1 Feature space and algorithm. Each asset i in the eligible universe is characterized by its mean daily log-return and return standard deviation over the 730-day window, both z-score standardized before clustering. We apply the CLARA algorithm of Kaufman and Rousseeuw (1990), which iterates over $Q = 50$ random subsamples and retains the partition that minimizes mean dissimilarity. We prefer CLARA to k-means because its medoid is an actual observed asset, which keeps it interpretable in economic terms and robust to outliers (Schubert and Rousseeuw, 2019).

3.4.2 Cluster number selection. The number of clusters k is set at each window by majority vote across four cluster validity indices, namely Silhouette (Rousseeuw, 1987), Dunn, Davies–Bouldin (Davies and Bouldin, 1979) and Caliński–Harabasz (Caliński and Harabasz, 1974), each evaluated over $k \in \{2, \dots, 15\}$. Clusters with fewer than 10 assets are excluded from the voting pool, following Lorenzo and Arroyo (2023), and ties are broken by the Silhouette coefficient.

3.4.3 Cluster selection. From each of the k^* valid clusters, we rank the members by their in-sample Sharpe ratio $\hat{\mu}_i/\hat{\sigma}_i$ and pass the top four to Stage 2. Pooling these across the valid

Table 2. Stage 0 tail-index filter: asset funnel across 75 windows

Filter stage	Mean assets	Range	% Of universe
Universe (pre-filter)	451	–	100.0%
Pass F1 (data coverage $\geq 90\%$)	443.6	[435, 451]	98.4%
Pass F2 (tail index $\geq$ median and ≥ 1.5)	222.1	[218, 226]	49.2%
Pass F3 (ann. vol $> 10\%$) — eligible set	222.1	[218, 226]	49.2%

Note(s): Mean and [min, max] across 75 walk-forward windows. F1: data coverage $\geq 90\%$. F2: tail index $\hat{\alpha}^{\min} \geq$ cross-sectional median and ≥ 1.5 . F3: annualized volatility $> 10\%$. F2 is the binding constraint

Table 3. Characteristics of stage 0 retained assets

Characteristic (retained assets)	Mean across windows	Range [min, max]
Ann. return	9.8%	[5.8%, 14.4%]
Ann. volatility	31.7%	[28.0%, 36.5%]
Tail index $\hat{\alpha}^{\min}$	2.529	[2.317, 2.686]
Adaptive F2 threshold	2.180	[1.690, 2.530]

Note(s): Per-asset statistics computed over the 730-day estimation window; means averaged across the 75 windows. Adaptive threshold = cross-sectional median of $\hat{\alpha}^{\min}$, the binding component of filter F2

clusters yields the investable set, which over the 75 evaluation windows comprises a mean of 22.4 assets (range: 8–44), with an average Silhouette coefficient of 0.377.

Figure 1 plots the evolution of k^* across the 75 windows: low- k partitions ($k \leq 5$) predominate during trending bull markets, whereas high- k partitions emerge under cross-sectional dispersion, notably during the 2022 bear market, confirming that the algorithm responds adaptively to market regimes.

3.5 Stage 2: CVaR-risk parity with GARCH-EVT

Stage 2 constructs a CVaR-risk-parity (CVaR-RP) portfolio over the N^* assets from Stage 1. The risk-parity framework follows Maillard *et al.* (2010) and Sheng *et al.* (2025), but it replaces their Gaussian CVaR specification with a three-step non-Gaussian tail model, which directly addresses the distributional limitation identified by Sheng *et al.* (2025).

3.5.1 GARCH(1,1) volatility modeling. For each asset $i = 1, \dots, N^*$, we fit a GARCH(1,1) model (Engle, 1982; Bollerslev, 1986) to the daily log-return series over the 730-day estimation window:

$$r_{it} = \mu_i + \varepsilon_{it}, \quad \varepsilon_{it} = \sigma_{it} z_{it}, \quad z_{it} \sim N(0, 1), \quad (2)$$

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2. \quad (3)$$

Estimation is by maximum likelihood. The fitted model yields standardized residuals that strip out conditional heteroskedasticity, together with a one-step-ahead volatility forecast used to scale the simulated shocks below. When GARCH fails to converge, we substitute an exponentially weighted moving average (span 30); in our sample, GARCH converges for every asset-window pair, so this fallback is never triggered.

3.5.2 GPD tail modeling. The standardized residuals $\hat{z}_{it}$ exhibit excess kurtosis and fat tails. Following McNeil and Frey (2000), we apply the Peaks-over-Threshold (POT) method and set the threshold u_i at the 5th percentile of $\hat{z}_{it}$. The Pickands–Balkema–de Haan theorem implies that left-tail exceedances follow a Generalized Pareto Distribution (GPD):

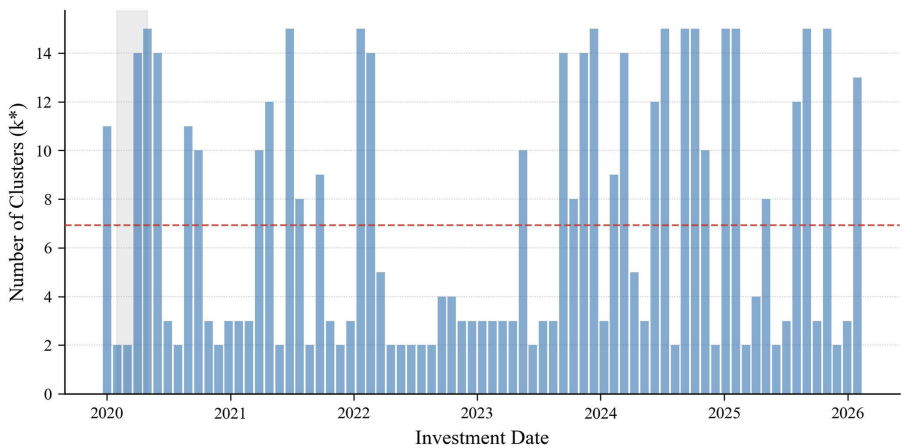


Figure 1. Number of clusters k^* selected by CLARA per investment window, January 2020 to January 2026. Dashed line: cross-window mean (6.9). Shaded band: COVID-19 disruption (February–April 2020)

$$P(-\hat{z}_i > u_i + y | -\hat{z}_i > u_i) \approx [1 + \xi_i y / \beta_i]^{-1/\xi_i}. \quad (4)$$

Parameters (ξ_i, β_i) are estimated by maximum likelihood on the tail exceedances. The resulting semi-parametric marginal CDF $\hat{F}_i(\cdot)$ uses the empirical distribution in the body and the GPD extension in the left tail, ensuring that CVaR estimates accurately reflect the heavy-tailed nature of equity returns (Artzner et al., 1999; Rockafellar and Uryasev, 2000).

3.5.3 Student-t copula and Monte Carlo simulation. We map each standardized residual to the uniform interval through its semi-parametric CDF: $U_{it} = \hat{F}_i(\hat{z}_{it}) \in (0, 1)$. We then fit a Student- t copula $C(u_1, \dots, u_n; R, \nu)$ by maximum likelihood over $\nu \in \{3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 30\}$. The t -copula generates asymptotic tail dependence for any finite ν , so extreme joint losses are more probable than the Gaussian copula admits (Embrechts et al., 2002). We simulate $M = 10,000$ joint return scenarios by inverting correlated uniform draws through each marginal CDF to recover standardized shocks $\hat{z}_i^m$, and then scale them to next-period returns $\tilde{r}_i^m = \hat{\mu}_i + \hat{\sigma}_{i,t+1} \cdot \hat{z}_i^m$.

3.5.4 CVaR-RP optimization. Given the M simulated loss scenarios $L^m = -\sum_i w_i \tilde{r}_i^m$, the 95% portfolio CVaR is:

$$\text{CVaR}_\alpha(w) = \frac{1}{\alpha M} \sum_{m: L^m \geq \text{VaR}_\alpha} L^m, \quad \alpha = 0.05. \quad (5)$$

The tail risk contribution of asset i is $\text{TRC}_i(w) = w_i \cdot \partial \text{CVaR}_\alpha(w) / \partial w_i$ (Tasche, 2002). CVaR-RP imposes equal contributions across assets, which we enforce by solving:

$$\min_w \sum_i \left(\frac{\text{TRC}_i(w)}{\text{CVaR}_\alpha(w)} - \frac{1}{N^*} \right)^2 \quad \text{s.t.} \quad \sum_i w_i = 1, \quad w_i \geq 0.01 \quad \forall i. \quad (6)$$

The lower bound $w_i \geq 0.01$ prevents degenerate corner solutions and enforces minimum diversification. Problem (6) is solved via Sequential Quadratic Programming (SLSQP). The resulting weights w^o constitute the Stage 2 allocation and are passed to Stage 3.

Figure 2 plots the fitted copula degrees of freedom across the 75 windows. The distribution is bimodal: low- ν windows ($\nu \leq 10$) account for 53% of the sample and coincide with elevated tail-contagion episodes, whereas calm expansionary periods produce ν in the range 20–30, confirming that the Student- t copula adapts dynamically to prevailing dependence regimes.

3.6 Stage 3: machine learning ensemble

CVaR-RP allocates capital on the basis of tail-risk equalization and takes no view on return direction. Stage 3 exploits residual predictability by training an eight-classifier ensemble to forecast 30-day return direction, then excluding predicted losers and reallocating their weight to predicted winners. This design adds return momentum without disturbing the tail-risk structure achieved in Stage 2.

3.6.1 Prediction target. For each asset i and investment date t , the binary label is $y_{it} = 1 \{P_{i,t+30} > P_{it}\}$, where P_{it} is the adjusted price on date t and $P_{i,t+30}$ is the price 30 calendar days forward.

3.6.2 Feature construction. Two complementary feature blocks are constructed from the 730-day estimation window. The tabular block comprises 17 features spanning multi-horizon momentum, return volatility, RSI, distributional moments and trend signals. The sequential block is the 60-day z-score-standardized return series, which preserves temporal ordering for sequence-sensitive classifiers; we validate this 60-day lookback against 120- and 240-day alternatives in the lookback sensitivity analysis of Section 4.

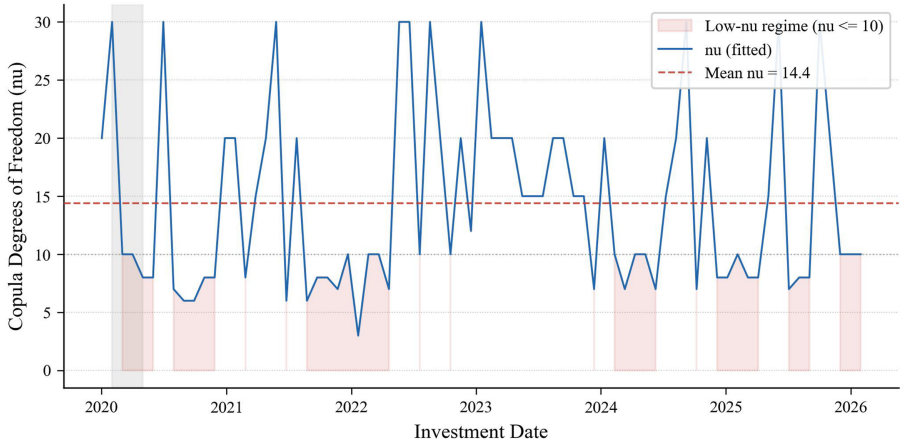


Figure 2. Fitted student- t copula degrees of freedom (ν) per investment window. Shaded area: low- ν regime ($\nu \leq 10$), associated with elevated tail contagion. Dashed line: cross-window mean (14.4). Shaded band: COVID-19 disruption

3.6.3 Classifier ensemble. Four tabular classifiers consume the 17-feature tabular input – Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosted Trees (GBDT) – and four sequence classifiers consume the 60-day sequential input: a CNN, an RNN, an LSTM and a Transformer. This architecturally diverse ensemble follows [Fischer and Krauss \(2018\)](#) and [Gu et al. \(2020\)](#) and guards against idiosyncratic model failure in a noisy prediction environment.

3.6.4 Walk-forward training. Each classifier is trained on all prior windows and generates predictions for the current window, which rules out look-ahead bias. We require a minimum of 12 prior windows before the first prediction, so the ensemble is active for 63 of the 75 evaluation windows. In the first 12 windows, we apply the Stage 2 weights directly.

3.6.5 Ensemble aggregation and weight adjustment. The ensemble prediction for asset i is formed by hard majority vote across the eight classifiers:

$$\hat{y}_{it} = 1 \text{ if } \sum_j \hat{y}_{it}^j \geq 4; \quad \hat{y}_{it} = 0 \text{ otherwise.} \quad (7)$$

Letting $S_t^+ = \{i : \hat{y}_{it} = 1\}$ denote the predicted winners, the Stage 3 weights are:

$$w_i^T = w_i^o / \sum_{j \in S_t^+} w_j^o \text{ if } i \in S_t^+; \quad w_i^T = 0 \text{ otherwise.} \quad (8)$$

Predicted losers receive zero weight; the freed capital is redistributed proportionally to predicted winners, preserving the relative Stage 2 CVaR-RP weights within the winning set. If $|S_t^+| < 2$, Stage 2 weights are retained without adjustment to prevent excessive portfolio concentration.

Figure 3 plots ensemble and per-model out-of-sample accuracy across the 63 active windows; ensemble accuracy does not degrade over time as the training set expands.

3.7 Evaluation framework

3.7.1 Benchmark strategies. We evaluate the proposed pipeline against 10 benchmarks, ordered from the simplest passive rule to the most complete optimizer so that each adds one

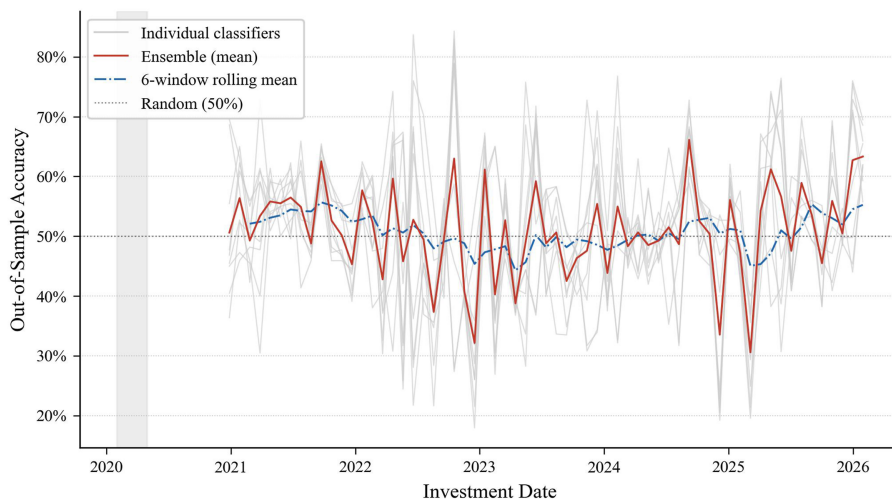


Figure 3. Out-of-sample accuracy of individual classifiers (gray) and ensemble mean (red) across 63 active walk-forward windows. Dash-dot line: 6-window rolling mean. Dotted line: random-chance baseline (50%)

modeling ingredient to the previous one; reading the list in order isolates the marginal value of market access, diversification, clustering, risk model and asset selection. Table 4 records which components each strategy activates and distinguishes the ten benchmarks from our two proposed strategies, CVaR_EVT and CVaR_ML.

The first benchmark, the S&P 500 ETF (SPY) held continuously, fixes the passive market return. Two naive diversification rules applied to the eligible universe follow: a market-capitalization-weighted portfolio (MCW), proxied by 90-day average dollar volume since we do not use shares-outstanding data and the equal-weighted 1/N rule (EW; DeMiguel *et al.*,

Table 4. Pipeline component activation across strategies

Strategy	CLARA clust	GARCH (1,1)	EVT/ GPD	<i>t</i> -copula	CVaR- RP optim	ML wt. adj
SPY B&H	X	X	X	X	X	X
MCW	X	X	X	X	X	X
EW	X	X	X	X	X	X
CL_EW	✓	X	X	X	X	X
MR_MV	✓	X	X	X	X	X
RP	✓	X	X	X	X	X
HRP	✓	X	X	X	X	X
CVaR_Normal	✓	X	X	X	✓	X
CVaR_EVT_TopN	X	✓	✓	✓	✓	X
CVaR_ML_TopN	X	✓	✓	✓	✓	✓
CVaR_EVT	✓	✓	✓	✓	✓	X
CVaR_ML ^a	✓	✓	✓	✓	✓	✓

Note(s) ^a Denotes the full proposed pipeline. RP, HRP and CVaR_Normal use the Stage 1 CLARA selection as their asset universe; the two TopN variants select the top-22 assets by in-sample Sharpe ratio and mirror CVaR_EVT and CVaR_ML, differing only in asset selection and thereby isolating the contribution of Stage 1 clustering. The modeling components are nested (CVaR_Normal → CVaR_EVT → CVaR_ML): CVaR_EVT adds the full EVT block (GARCH(1,1) filtering, GPD tail modeling and *t*-copula dependence) and CVaR_ML further adds the ML weight adjustment

2009). The two strategies of [Lorenzo and Arroyo \(2023\)](#) form the baseline from which our method departs: CL_EW equal-weights the assets of the valid CLARA clusters, and MR_MV applies global minimum-variance optimization over the cluster prototypes, both using Stage 1 clustering but no tail modeling. Holding the universe fixed at the Stage 1 selection, three further benchmarks vary only the risk model and so isolate our GARCH-EVT-copula contribution: risk parity (RP; [Maillard et al., 2010](#)), HRP ([López de Prado, 2016](#)) and Gaussian CVaR-RP (CVaR_Normal; [Sheng et al., 2025](#)), where the step from CVaR_Normal to CVaR_EVT measures the value of replacing the Gaussian assumption with the full EVT tail model. Finally, CVaR_EVT_TopN and CVaR_ML_TopN replace CLARA with a top-22 in-sample-Sharpe ranking while leaving the rest of the pipeline unchanged, so comparing each proposed strategy with its TopN counterpart isolates the contribution of cluster-based pre-selection.

Against these ten benchmarks, we set our two proposed strategies. CVaR_EVT applies the full Stage 2 optimizer, layering GARCH(1,1) volatility filtering, GPD tail modeling and t -copula dependence onto the CLARA universe. CVaR_ML adds the Stage 3 machine-learning weight adjustment on top.

3.7.2 Performance metrics. We report annualized return and volatility, the Sharpe, Sortino and Calmar ratios, maximum drawdown, monthly win rate, realized 95% CVaR and mean one-way monthly turnover across the 75 out-of-sample windows, net of 5 bps transaction costs. We set the risk-free rate to zero for cross-strategy comparability.

3.7.3 Bootstrap inference. We supplement all results with block-bootstrap inference following [Politis and Romano \(1994\)](#), drawing $B = 1,000$ resamples of the 75-month return matrix with 6-month moving blocks. We report 90% bootstrap confidence intervals and bootstrap mean Sharpe ratios, a less path-dependent measure of strategy quality than single realized estimates ([Kaur et al., 2026](#); [Silva and Andrade, 2024](#)).

4. Results

4.1 Pipeline diagnostics

We begin by examining how each stage behaves over the sample. The Hill filter opens the process: on average it screens 444 constituents per window and keeps 222.1 eligible assets, a retention rate of 50.1% (range: 218 to 226). The adaptive threshold moves between 1.69 and 2.53 (mean 2.18), falling to its sample minimum during the COVID-19 shock of April to July 2020, when return distributions grow heavier tails across the whole universe at once; the filter therefore tightens exactly when tail risk is highest. Filter F3, the minimum-volatility screen, adds little on the S&P 500 universe. [Figure 4](#) traces how the number of eligible tickers evolves across the 75 windows.

The CLARA algorithm then works on the roughly 222 eligible assets in each window, selecting k endogenously between 2 and 15 (median 3; mean 6.9). Low- k partitions ($k \leq 5$) dominate in trending bull markets, whereas high- k partitions appear when the cross-section disperses, most clearly during the 2022 bear market. The mean Silhouette coefficient is 0.377 (range 0.310–0.479), a moderate separation consistent with the known weakness of return-space clustering in large equity universes. From each valid cluster, the four highest-Sharpe assets advance, yielding a mean investable set of 22.4 assets (range 8–44).

At the optimization stage, the fitted Student- t copula degrees of freedom range from 3 to 30 across windows (mean 14.4). Low values ($\nu \leq 10$, 53% of windows) cluster in periods of market stress, notably the COVID-19 shock and the 2022 bear market, while expansionary windows produce ν between 20 and 30; because a lower ν implies stronger tail dependence in elliptical copulas ([Embrechts et al., 2002](#)), this variation gives a direct, data-driven measure of tail contagion. Portfolio CVaR₉₅ averages 4.24% across windows (range 2.91%–11.45%), peaking during the COVID-19 shock. The CVaR/VaR ratio averages 1.416, well above the Gaussian benchmark of 1.24, confirming that the framework captures heavier joint tails than the Gaussian assumption allows ([McNeil and Frey, 2000](#)). The risk-parity constraint holds

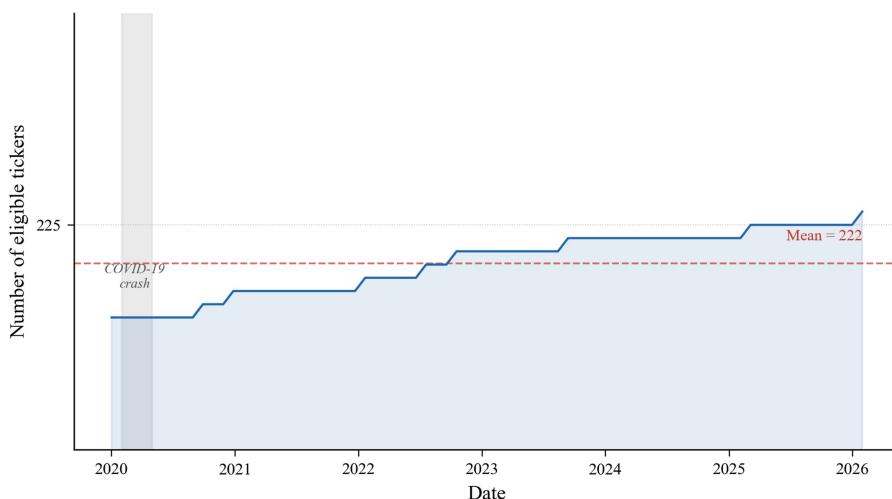


Figure 4. Number of S&P 500 tickers passing the EVT tail-risk filter per investment window, January 2020 to January 2026. The filter retains stocks whose Hill tail-index estimate exceeds the cross-sectional median threshold. The shaded band marks the COVID-19 market disruption (February–April 2020)

with near-perfect precision (mean within-window standard deviation of CVaR contributions of only 0.01%).

The machine-learning layer comes last. The eight-classifier ensemble reaches a mean out-of-sample accuracy of 52.6% across 63 walk-forward windows (standard deviation 13.0%). GBDT is the strongest single model, leading on accuracy in 29% of windows and on F1 in 57%, consistent with the known edge of gradient-boosted trees on low-signal financial tabular data (Gu *et al.*, 2020; Krauss *et al.*, 2017). This accuracy sits within the broader directional-forecasting literature, where Krauss *et al.* (2017), Fischer and Krauss (2018) and Grądzki *et al.* (2025) report out-of-sample accuracies around 50%–56% and find that the most accurate model need not deliver the highest economic return. Accuracy also holds up over time, rising from 54.4% to 55.6% between the first and last ten windows as the training set expands, and a lookback sensitivity analysis confirms the 60-day sequential window as the best specification (Brier score 0.257, against 0.260 and 0.263 for the 120- and 240-day windows). Following Gu *et al.* (2020), we therefore judge the ML layer mainly by its effect on the portfolio rather than by classification accuracy alone.

The ML layer adjusts portfolio weights modestly. The mean weight change from Stage 2 to Stage 3 is + 1.3% points (standard deviation: 2.4%), and the correlation between Stage 2 and Stage 3 weights is 0.889. We read these figures as a measured tactical tilt rather than a wholesale reallocation. Figure 5 summarizes per-model classification performance.

4.2 Portfolio performance

We now move from pipeline diagnostics to the question that matters for investors: out-of-sample performance. Table 5 reports the results of twelve strategies across the 75 monthly windows (January 2020 to January 2026), net of 5 basis points of transaction costs per unit of one-way turnover. The strategies range from passive and naïve benchmarks (SPY B&H, MCW, EW) through cluster- and risk-based comparators to our proposed four-stage pipeline (CVaR_EVT and CVaR_ML).

Our first comparison asks whether EVT tail estimation protects the downside better than Gaussian CVaR. CVaR_EVT and CVaR_Normal share the same CLARA universe and

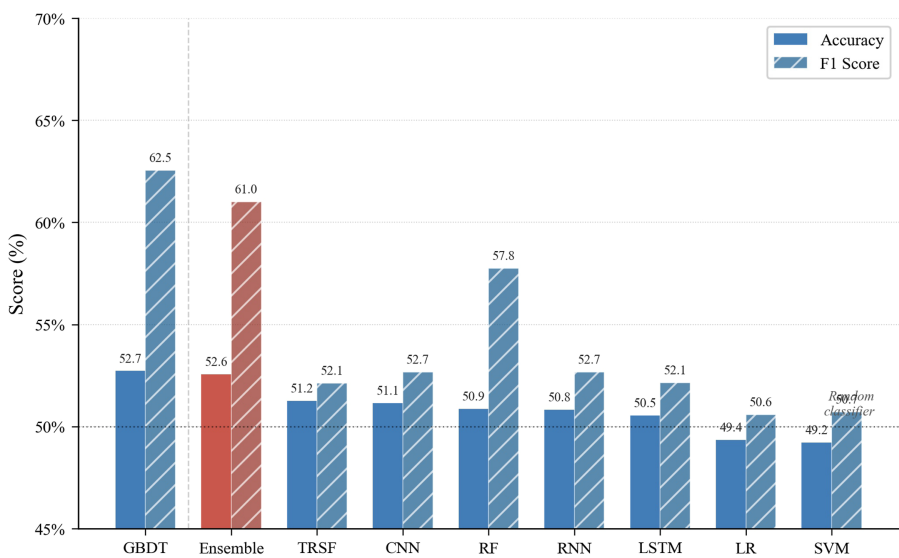


Figure 5. Out-of-sample classification performance of base learners and the ensemble, averaged across all rolling investment windows. LR = Logistic Regression; SVM = Support Vector Machine; RF = Random Forest; GBDT = Gradient-Boosted Decision Trees; CNN = Convolutional Neural Network; RNN = Recurrent Neural Network; LSTM = Long Short-Term Memory; TRSF = Transformer

Table 5. Out-of-sample portfolio performance across twelve strategies (75 monthly windows, January 2020–January 2026). All metrics net of 5 bps transaction costs. *Italic row: CVaR_ML (proposed model)*

Strategy	Ann. Ret	Ann. Vol	Sharpe	Sortino	MaxDD	Calmar	CVaR ₉₅	Win rate	Turnover
SPY B&H	14.3%	18.4%	0.776	0.942	−23.0%	0.620	−12.7%	64.0%	0.00
MCW	18.8%	25.9%	0.729	1.099	−30.0%	0.628	−14.1%	62.7%	0.34
EW	15.0%	21.6%	0.694	0.783	−31.7%	0.473	−14.4%	64.0%	0.25
CL_EW	13.5%	21.0%	0.643	0.717	−31.7%	0.427	−14.4%	61.3%	0.35
MR_MV	18.4%	22.0%	0.836	0.890	−27.2%	0.676	−14.9%	69.3%	1.61
RP	26.7%	24.5%	1.089	1.582	−28.5%	0.936	−13.8%	69.3%	1.67
HRP	25.6%	24.0%	1.064	1.471	−27.1%	0.943	−14.2%	68.0%	1.65
CVaR_EVT_TopN	13.6%	20.9%	0.649	0.738	−28.7%	0.473	−13.8%	68.0%	0.74
CVaR_ML_TopN	14.8%	21.4%	0.695	0.901	−23.0%	0.646	−15.6%	64.0%	1.00
CVaR_Normal	26.5%	24.5%	1.080	1.581	−28.6%	0.926	−14.1%	66.7%	1.63
CVaR_EVT	20.9%	21.1%	0.988	0.961	−27.4%	0.762	−14.4%	74.7%	1.48
CVaR_ML	24.1%	21.8%	1.106	1.118	−27.4%	0.880	−14.6%	76.0%	1.62

Note(s): Ann. Ret = annualized return; Ann. Vol = annualized volatility; MaxDD = maximum drawdown; Calmar = Ann. Ret/|MaxDD|; Win Rate = fraction of months with positive net return; Turnover = average monthly one-way portfolio turnover. The risk-free rate is set to zero for cross-strategy comparability. CL_EW: equal-weighted portfolio over all assets in valid CLARA clusters. MR_MV: global minimum-variance over the cluster prototypes. CVaR_Normal: CVaR-RP with Gaussian marginals, following [Sheng et al. \(2025\)](#). CVaR_EVT: Stage 0–2 pipeline without ML adjustment. CVaR_ML: full four-stage pipeline (proposed). CVaR₉₅ (rlz.) = realized 95% CVaR computed on the 75 out-of-sample monthly returns (annualized)

CVaR-risk-parity objective but differ across the entire risk model: CVaR_Normal uses static Gaussian marginals with a closed-form Gaussian CVaR, whereas CVaR_EVT adds GARCH(1,1) filtering, GPD tail modeling and a Student-*t* copula, so the contrast isolates the value of the full non-Gaussian tail machinery. We find a consistent EVT advantage in capital preservation: CVaR_EVT cuts the maximum drawdown from -28.6% to -27.4% and lifts the monthly win rate from 66.7% to 74.7% . This protection has a cost in raw return, as CVaR_EVT earns 20.9% annualized against 26.5% for CVaR_Normal, with a lower Sharpe ratio (0.988 versus 1.080). This is the intended effect of the specification: by perceiving heavier joint tails accurately, the optimizer diversifies more aggressively (Rockafellar and Uryasev, 2000), giving up some upside in the bull-market windows that dominate our sample while preserving capital better during stress, in line with the finding that tail dependence rises sharply in down markets (Longin and Solnik, 2001). Realized CVaR₉₅ shows that EVT-constrained strategies do not lower tail losses in absolute terms (SPY B&H records the lowest realized CVaR, 12.7% annualized); instead, they reach higher Sharpe ratios by earning more return per unit of risk, consistent with evidence that GARCH-EVT-Copula models improve out-of-sample performance relative to historical-distribution benchmarks (Sahamkhadam *et al.*, 2018).

Our second comparison isolates the contribution of the ML ensemble by setting CVaR_ML against CVaR_EVT. Adding the Stage 3 overlay raises the Sharpe ratio (1.106 versus 0.988), the annualized return (24.1% versus 20.9%), the Calmar ratio (0.880 versus 0.762) and the win rate (76.0% versus 74.7%), while the maximum drawdown stays at -27.4% , so the overlay does not weaken the tail-risk discipline of Stage 2. Stage 3 is therefore an efficient contribution rather than a marginal one: a modest tactical tilt (a mean weight change of only 1.3% points) delivers a 12% improvement in the Sharpe ratio at no cost to downside protection, consistent with recent evidence that combining GARCH-EVT-Copula with deep-learning models improves risk-adjusted performance over optimization-only baselines (Huynh and Ta, 2025).

Our third comparison isolates the contribution of CLARA clustering by setting each proposed strategy against its TopN counterpart, which replaces cluster-based selection with the top 22 assets ranked by in-sample Sharpe ratio while leaving the rest of the pipeline unchanged. The effect is decisive: removing clustering cuts the Sharpe ratio of CVaR_EVT from 0.988 to 0.649 and that of CVaR_ML from 1.106 to 0.695 , and lowers the monthly win rate in both cases (74.7% – 68.0% and 76.0% – 64.0%). Both TopN variants fall below the passive SPY benchmark (0.776), and the block-bootstrap means confirm the collapse (0.769 and 0.784 against 1.022 for SPY). Cluster-based pre-selection is therefore a structural precondition for the pipeline rather than a computational convenience.

The performance ranking confirms that CVaR_ML leads on risk-adjusted metrics, ranking first among all twelve strategies on the point-estimate Sharpe ratio (1.106) and the monthly win rate (76.0%); CVaR_EVT ties it for the lowest maximum drawdown (-27.4%) among EVT-constrained strategies. Against SPY B&H, CVaR_ML adds 9.8% points of annualized return (24.1 versus 14.3%) while its maximum drawdown is only 4.4% points deeper (-27.4% versus -23.0%), and its cumulative return of 285.4% far exceeds both SPY B&H (130.4%) and CL_EW (120.9%).

Not every metric favors CVaR_ML. Figure 6 shows that it does not produce the highest cumulative wealth in the sample, since RP accumulates 4.38 times the initial investment against 3.85 times for CVaR_ML. This gap is a structural consequence of the tail-risk constraint rather than an episodic artifact: RP and HRP impose no accurate tail-risk constraint (Maillard *et al.*, 2010; Lohre *et al.*, 2020), yet their drawdowns are comparable to ours (-28.5% for RP and -27.1% for HRP, against -27.4% for CVaR_ML), so the unconstrained strategies do not buy their higher cumulative returns through better downside control but through higher returns earned at higher volatility. The post-COVID recovery of 2020–2021 and the AI-driven bull market of 2023–2025 are exactly the regimes where accurate tail constraints cut weight in high-momentum assets the most, and where the cumulative gap is widest. On the Calmar ratio (CVaR_ML 0.880 ; RP 0.936 ; HRP 0.943), RP and HRP rank

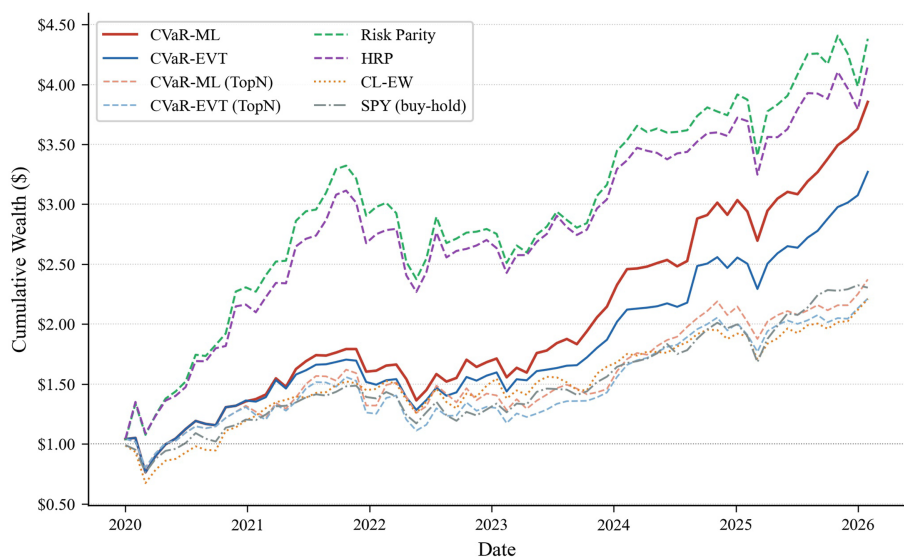


Figure 6. Cumulative wealth of selected portfolio strategies, January 2020 to January 2026. Each series is initialized at \$1.00 in January 2020 and compounds monthly net returns. SPY_BH denotes a buy-and-hold investment in the S&P 500 ETF

ahead because their higher returns offset slightly deeper drawdowns; CVaR_ML’s advantage is one of risk-adjusted efficiency and consistency rather than terminal wealth.

Figure 7 presents the drawdown paths of the selected strategies over the evaluation period, complementing the cumulative-wealth comparison in Figures 6 and 8 plots the 12-month

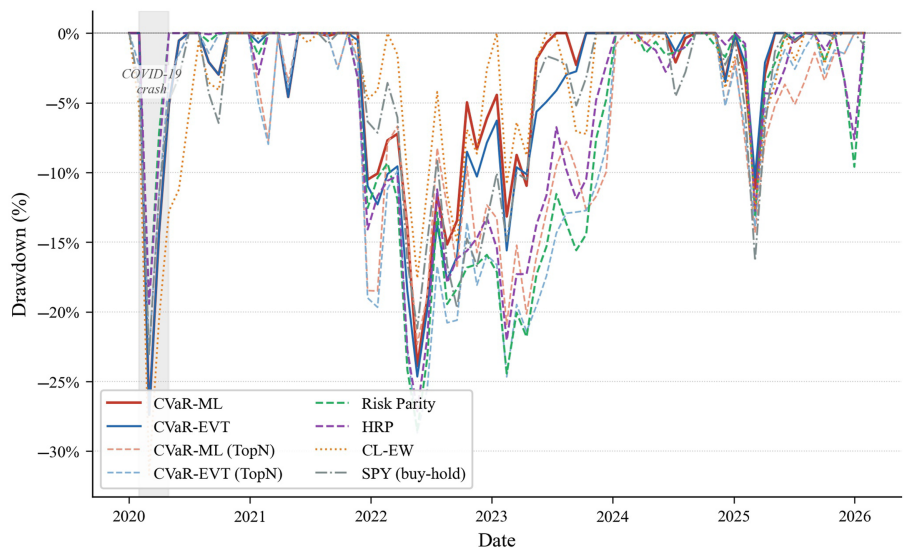


Figure 7. Drawdown paths for selected portfolio strategies, January 2020 to January 2026. The shaded band marks the COVID-19 market disruption (February–April 2020)

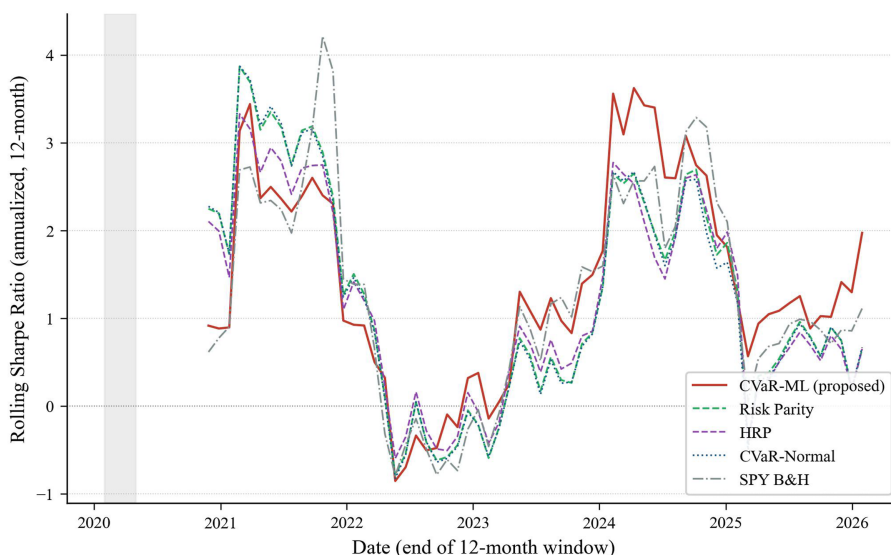


Figure 8. Rolling 12-month annualized Sharpe ratio for selected strategies, January 2021 to January 2026. Shaded band: COVID-19 disruption (February–April 2020)

rolling Sharpe ratio. CVaR_{ML} leads across regimes, except during the 2020 to 2021 momentum recovery, when unconstrained strategies gain from concentrated positions that our tail-risk budget rules out. Its rolling advantage is clearest in the 2022 bear market and the 2023 to 2024 consolidation. We take this pattern as evidence that the outperformance of CVaR_{ML} is not limited to a single episode.

4.3 Bootstrap robustness

Point estimates based on 75 monthly observations depend on the path of returns. To address this, we apply block bootstrap inference following Politis and Romano (1994) and Ledoit and Wolf (2008). We draw $B = 1,000$ 6-month moving-block resamples and re-evaluate every strategy on each one. Table 6 reports the bootstrap mean Sharpe ratios, maximum drawdowns and cumulative returns, together with 90% confidence intervals for selected strategies.

Figure 9 displays the bootstrap mean Sharpe ratios and their 90% intervals across strategies. Three findings emerge. The bootstrap means exceed the single-path realized Sharpe ratios for every strategy, a consequence of resampling short blocks with replacement and of the nonlinearity of the Sharpe ratio, so we read them as a less path-dependent ranking device rather than as point forecasts. First, CVaR_{ML} records the highest bootstrap mean Sharpe ratio (1.316), ahead of CVaR_{EVT} (1.188), RP (1.183), CVaR_{Normal} (1.166) and HRP (1.116), so the point-estimate ranking survives resampling. Second, the 90% confidence intervals for the top active strategies are wide and overlap substantially, given the 75-observation sample, so individual rankings cannot be asserted with high precision; the decisive comparison is that the P5 Sharpe of CVaR_{ML} (0.534) exceeds that of SPY B&H (0.306), confirming a statistically robust advantage over the passive benchmark. Third, the fifth-percentile Sharpe for CL_{EW} (0.211) stays below that of SPY B&H, so the equal-weighted filtered strategy does not offer robust Sharpe outperformance.

Table 6. Bootstrap inference ($B = 1,000$ block-bootstrap resamples, 6-month moving blocks, 90% confidence intervals). *Italic row: CVaR_ML*

Strategy	Sharpe (mean)	Sharpe (90% CI)	MaxDD (mean)	MaxDD (90% CI)	CumRet (mean)	CumRet (90% CI)
SPY B&H	1.022	[0.306, 1.878]	−23.2%	[−35.6%, −11.7%]	179%	[44%, 352%]
RP	1.183	[0.320, 2.120]	−24.3%	[−40.4%, −13.1%]	377%	[52%, 972%]
HRP	1.116	[0.384, 1.939]	−23.2%	[−37.7%, −13.2%]	318%	[67%, 726%]
CVaR_Normal	1.166	[0.341, 2.074]	−23.8%	[−39.1%, −13.0%]	367%	[54%, 918%]
CVaR_EVT	1.188	[0.449, 2.104]	−24.0%	[−36.7%, −10.5%]	273%	[73%, 578%]
<i>CVaR_ML</i>	<i>1.316</i>	<i>[0.534, 2.228]</i>	−23.9%	[−35.7%, −11.2%]	349%	[96%, 727%]
CVaR_EVT_TopN	0.769	[0.060, 1.638]	−27.8%	[−44.1%, −14.6%]	150%	[9%, 349%]
CVaR_ML_TopN	0.784	[0.119, 1.603]	−25.1%	[−39.1%, −16.3%]	162%	[18%, 367%]

Note(s): Bootstrap mean and 90% confidence intervals (CI; P5-P95) from 1,000 block-bootstrap replications with 6-month moving blocks. Cumulative returns expressed as percentage gains on the initial investment

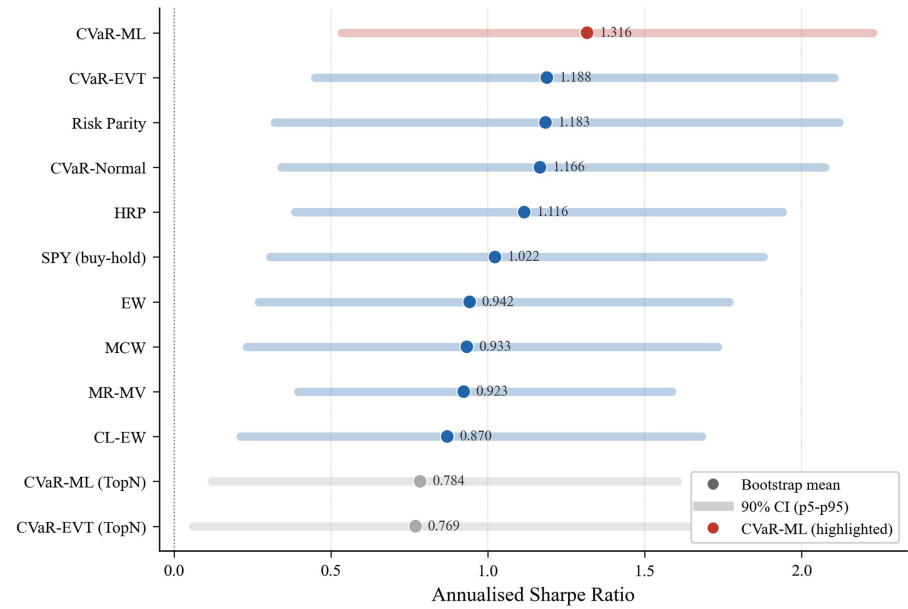


Figure 9. Bootstrap distribution of annualized Sharpe ratios across strategies. Points denote bootstrap means; horizontal bars span the 5th to 95th percentiles from 1,000 block-bootstrap replications with six-month moving blocks. Strategies are sorted by bootstrap mean Sharpe ratio in descending order

4.4 Turnover and transaction cost sensitivity

Finally, we test how sensitive the results are to turnover and trading costs. The average monthly one-way turnover of CVaR_ML is 1.62, close to RP (1.67) and HRP (1.65) and well above CL_EW (0.35), which rebalances little (DeMiguel *et al.*, 2009). At 5 bps per unit of turnover, the monthly cost of CVaR_ML is about 8 bps, or roughly 97 bps a year. Under cost scenarios of 5, 10 and 20 bps per unit of monthly turnover, CVaR_ML keeps the highest net Sharpe ratio in every case, so the ranking holds under different cost assumptions. These results fit evidence

that real transaction costs for institutional investors in large-cap equity markets are well below common assumptions (Frazzini *et al.*, 2012) and that strategies with monthly one-sided turnover well under 50% still earn significant net spreads after costs (Novy-Marx and Velikov, 2016). Trading costs therefore reduce, but do not remove, the risk-adjusted advantage of CVaR_ML for institutional investors with standard large-cap access.

5. Conclusion

This article proposes and tests a four-stage portfolio construction pipeline designed to address three concurrent failure modes of standard Gaussian frameworks: dimensionality in large equity universes, distributional misspecification in tail-risk estimation and the discarding of forward-looking return information. We implement it on S&P 500 constituents across 75 monthly walk-forward windows spanning January 2020 to January 2026, a period covering the COVID-19 collapse, the 2022 rate-driven bear market and two extended bull phases. The pipeline applies a rolling Hill estimator to screen for tail-eligible assets, clusters the survivors through CLARA and selects the highest-Sharpe cohort, builds a CVaR risk-parity portfolio from GARCH(1,1)-EVT marginals and a Student-*t* copula, and adds a tactical machine-learning overlay through an eight-classifier ensemble that operates strictly within the optimization-stage risk budget. In doing so, it supplies the forward-looking predictive signal that Lorenzo and Arroyo (2023) identified as missing and replaces the Gaussian assumption that Sheng *et al.* (2025) flagged for future work.

The central finding is best understood through what our strategy does not maximize: cumulative wealth. RP accumulates 4.38 times the initial investment over the sample, versus 3.85 times for CVaR_ML. This gap is structural: lacking an accurate tail-risk constraint, RP and HRP concentrate their additional return in the post-COVID recovery (2020–2021) and the 2023–2025 AI-driven bull market, precisely the regimes in which a binding tail budget pulls weight away from high-momentum names. CVaR_ML targets risk-adjusted returns under drawdown discipline, not unconditional wealth and on that criterion it succeeds: it ranks first among all 12 strategies on Sharpe ratio (1.106) and monthly win rate (76.0%). Bootstrap inference with 1,000 block replications confirms the ranking is not an artefact of path dependence, as CVaR_ML achieves the highest bootstrap mean Sharpe ratio (1.316) and its fifth-percentile Sharpe (0.534) exceeds that of passive SPY (0.306). The EVT constraint does not minimize realized CVaR in absolute terms (SPY B&H records the lowest, 12.7% annualized); the framework is a maximum-efficiency device operating under a tail-risk budget, not a minimum-loss device.

Three controlled comparisons support the design. Replacing Gaussian CVaR with EVT-based tail estimation reduces maximum drawdown from –28.6% to –27.4% and raises the monthly win rate from 66.7% to 74.7%, so accurate distributional assumptions improve capital preservation during stress. The machine-learning ensemble then lifts the Sharpe ratio from 0.988 to 1.106 and the Calmar ratio from 0.762 to 0.880 without widening drawdown, adding return at no cost to tail discipline. Most consequentially, removing the CLARA pre-selection collapses the Sharpe ratio to 0.649, below the passive benchmark, confirming that cluster-based pre-selection is a structural precondition rather than a computational convenience. Operationally, the pipeline’s annual transaction-cost load is roughly 97 basis points and CVaR_ML retains the highest net Sharpe ratio even under more conservative cost assumptions, confirming feasibility for institutional investors with standard large-cap access.

The study has several limitations. The pipeline is calibrated and evaluated exclusively on S&P 500 constituents; its behavior in smaller or less liquid universes, where the Hill estimator may be less reliable and cluster separation less stable, remains untested. With 75 walk-forward windows, point-estimate rankings carry wide confidence intervals. The ensemble’s mean out-of-sample accuracy of 52.6% is a modest signal that leaves room for richer feature engineering or alternative architectures. Future research could extend the framework to international

developed markets, incorporate dynamic copula specifications that let tail dependence evolve within the estimation window and explore whether option-implied volatility surfaces or analyst forecast revisions improve directional accuracy, potentially narrowing the cumulative-return gap relative to unconstrained risk parity without sacrificing tail-risk discipline.

Declaration on the use of AI tools

The manuscript was originally written in Spanish by the authors. A generative AI tool (Anthropic Claude, Opus 4.8, 2026) was used to translate the text into English and to assist with language and length editing to meet the journal's word limit. No research data, analyses, results or original arguments were generated by AI tools; all intellectual content and conclusions are the authors' own. The authors reviewed and verified all translated and edited text, including all numbers, figures, tables, citations and references, and take full responsibility for the accuracy and integrity of the work.

References

- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999), "Coherent measures of risk", *Mathematical Finance*, Vol. 9 No. 3, pp. 203-228, doi: [10.1111/1467-9965.00068](https://doi.org/10.1111/1467-9965.00068).
- Balkema, A.A. and de Haan, L. (1974), "Residual life time at great age", *Annals of Probability*, Vol. 2 No. 5, pp. 792-804, doi: [10.1214/aop/1176996548](https://doi.org/10.1214/aop/1176996548).
- Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, Vol. 31 No. 3, pp. 307-327, doi: [10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).
- Bradrania, R., Pirayesh Neghab, D. and Shafizadeh, M. (2022), "State-dependent stock selection in index tracking: a machine learning approach", *Financial Markets and Portfolio Management*, Vol. 36, pp. 1-28, doi: [10.1007/s11408-021-00391-7](https://doi.org/10.1007/s11408-021-00391-7).
- Bryzgalova, S., Pelger, M. and Zhu, J. (2025), "Forest through the trees: building cross-sections of stock returns", *The Journal of Finance*, Vol. 80 No. 5, pp. 2447-2506, doi: [10.1111/jofi.13477](https://doi.org/10.1111/jofi.13477).
- Caliński, T. and Harabasz, J. (1974), "A dendrite method for cluster analysis", *Communications in Statistics - Theory and Methods*, Vol. 3, pp. 1-27, doi: [10.1080/03610927408827101](https://doi.org/10.1080/03610927408827101).
- Cirulli, A., Kobak, M. and Ulrych, U. (2024), "Portfolio construction with hierarchical momentum", *The Journal of Portfolio Management*, Vol. 50 No. 4, pp. 159-175, doi: [10.3905/jpm.2023.1.570](https://doi.org/10.3905/jpm.2023.1.570).
- Davies, D.L. and Bouldin, D.W. (1979), "A cluster separation measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 1 No. 2, pp. 224-227, doi: [10.1109/TPAMI.1979.4766909](https://doi.org/10.1109/TPAMI.1979.4766909).
- DeMiguel, V., Garlappi, L. and Uppal, R. (2009), "Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy?", *The Review of Financial Studies*, Vol. 22 No. 5, pp. 1915-1953, doi: [10.1093/rfs/hhm075](https://doi.org/10.1093/rfs/hhm075).
- DeMiguel, V., Martín-Utrera, A., Nogales, F.J. and Uppal, R. (2020), "A transaction-cost perspective on the multitude of firm characteristics", *The Review of Financial Studies*, Vol. 33 No. 5, pp. 2180-2222, doi: [10.1093/rfs/hhz085](https://doi.org/10.1093/rfs/hhz085).
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modeling Extremal Events for Insurance and Finance*, Springer, Berlin.
- Embrechts, P., McNeil, A.J. and Straumann, D. (2002), "Correlation and dependence in risk management: properties and pitfalls", in Dempster, M.A.H. (Ed.), *Risk Management: Value at Risk and beyond*, Cambridge University Press, Cambridge, pp. 176-223, doi: [10.1017/CBO9780511615337.008](https://doi.org/10.1017/CBO9780511615337.008).
- Engle, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation", *Econometrica*, Vol. 50 No. 4, pp. 987-1008, doi: [10.2307/1912773](https://doi.org/10.2307/1912773).
- Fama, E.F. (1965), "The behavior of stock market prices", *The Journal of Business*, Vol. 38 No. 1, pp. 34-105, doi: [10.1086/294743](https://doi.org/10.1086/294743).

- Fischer, T. and Krauss, C. (2018), "Deep learning with long short-term memory networks for financial market predictions", *European Journal of Operational Research*, Vol. 270 No. 2, pp. 654-669, doi: [10.1016/j.ejor.2017.11.054](https://doi.org/10.1016/j.ejor.2017.11.054).
- Frazzini, A., Israel, R. and Moskowitz, T.J. (2012), "Trading costs of asset pricing anomalies", Fama-Miller Working Paper, Chicago Booth Research Paper No, pp. 14-05, doi: [10.2139/ssrn.2294498](https://doi.org/10.2139/ssrn.2294498).
- Grądzki, P., Wójcik, P. and Lessmann, S. (2025), "Algorithmic crypto trading using information-driven bars, triple barrier labeling and deep learning", *Financial Innovation*, Vol. 11 No. 1, p. 136, doi: [10.1186/s40854-025-00866-w](https://doi.org/10.1186/s40854-025-00866-w).
- Gu, S., Kelly, B. and Xiu, D. (2020), "Empirical asset pricing via machine learning", *The Review of Financial Studies*, Vol. 33 No. 5, pp. 2223-2273, doi: [10.1093/rfs/hhaa009](https://doi.org/10.1093/rfs/hhaa009).
- Hill, B.M. (1975), "A simple general approach to inference about the tail of a distribution", *The Annals of Statistics*, Vol. 3 No. 5, pp. 1163-1174, doi: [10.1214/aos/1176343247](https://doi.org/10.1214/aos/1176343247).
- Htun, H.H., Biehl, M. and Petkov, N. (2024), "Forecasting relative returns for S&P 500 stocks using machine learning", *Financial Innovation*, Vol. 10 No. 1, p. 118, doi: [10.1186/s40854-024-00644-0](https://doi.org/10.1186/s40854-024-00644-0).
- Huynh, V. and Ta, B.Q. (2025), "Black-Litterman portfolio optimization based on GARCH-EVT-Copula and LSTM models", *Annals of Operations Research*, Vol. 349 No. 3, pp. 1693-1715, doi: [10.1007/s10479-025-06597-6](https://doi.org/10.1007/s10479-025-06597-6).
- Kaufman, L. and Rousseeuw, P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- Kaur, S., Singh, A. and Aggarwal, A. (2026), "Optimal portfolio construction with fuzzy least square support vector machines and conditional value-at-risk: a risk-adjusted approach", *International Journal of System Assurance Engineering and Management*, Vol. 17 No. 4, pp. 1072-1106, doi: [10.1007/s13198-025-03020-y](https://doi.org/10.1007/s13198-025-03020-y).
- Kelly, B., Malamud, S. and Zhou, K. (2024), "The virtue of complexity in return prediction", *The Journal of Finance*, Vol. 79 No. 1, pp. 459-503, doi: [10.1111/jofi.13298](https://doi.org/10.1111/jofi.13298).
- Krauss, C., Do, X.A. and Huck, N. (2017), "Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500", *European Journal of Operational Research*, Vol. 259 No. 2, pp. 689-702, doi: [10.1016/j.ejor.2016.10.031](https://doi.org/10.1016/j.ejor.2016.10.031).
- Ledoit, O. and Wolf, M. (2008), "Robust performance hypothesis testing with the Sharpe ratio", *Journal of Empirical Finance*, Vol. 15 No. 5, pp. 850-859, doi: [10.1016/j.jempfin.2008.03.002](https://doi.org/10.1016/j.jempfin.2008.03.002).
- Lee, M.-C. (2025), "A hybrid EGARCH-Informer model with consistent risk calibration for volatility and CVaR forecasting", *Mathematics*, Vol. 13 No. 19, p. 3108, doi: [10.3390/math13193108](https://doi.org/10.3390/math13193108).
- Leippold, M., Wang, Q. and Zhou, W. (2022), "Machine learning in the Chinese stock market", *Journal of Financial Economics*, Vol. 145 No. 1, pp. 64-82, doi: [10.1016/j.jfineco.2021.11.017](https://doi.org/10.1016/j.jfineco.2021.11.017).
- Lohre, H., Rother, C. and Schäfer, K.A. (2020), "Hierarchical risk parity: accounting for tail dependencies in multi-asset multi-factor allocations", in Jurczenko, E. (Ed.), *Machine Learning for Asset Management: New Developments and Financial Applications*, Wiley, Hoboken, doi: [10.1002/9781119751182.ch9](https://doi.org/10.1002/9781119751182.ch9).
- Longin, F. and Solnik, B. (2001), "Extreme correlation of international equity markets", *The Journal of Finance*, Vol. 56 No. 2, pp. 649-676, doi: [10.1111/0022-1082.00340](https://doi.org/10.1111/0022-1082.00340).
- López de Prado, M. (2016), "Building diversified portfolios that outperform out of sample", *The Journal of Portfolio Management*, Vol. 42 No. 4, pp. 59-69, doi: [10.3905/jpm.2016.42.4.059](https://doi.org/10.3905/jpm.2016.42.4.059).
- Lorenzo, L. and Arroyo, J. (2023), "Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm", *Financial Innovation*, Vol. 9 No. 1, p. 25, doi: [10.1186/s40854-022-00438-2](https://doi.org/10.1186/s40854-022-00438-2).
- Maillard, S., Roncalli, T. and Teiletche, J. (2010), "The properties of equally weighted risk contribution portfolios", *The Journal of Portfolio Management*, Vol. 36 No. 4, pp. 60-70, doi: [10.3905/jpm.2010.36.4.060](https://doi.org/10.3905/jpm.2010.36.4.060).

- Markowitz, H. (1952), "Portfolio selection", *The Journal of Finance*, Vol. 7 No. 1, pp. 77-91, doi: [10.1111/j.1540-6261.1952.tb01525.x](https://doi.org/10.1111/j.1540-6261.1952.tb01525.x).
- Mba, J.C. and Mwambi, S. (2020), "A Markov-switching COGARCH approach to cryptocurrency portfolio selection and optimization", *Financial Markets and Portfolio Management*, Vol. 34 No. 2, pp. 199-214, doi: [10.1007/s11408-020-00346-4](https://doi.org/10.1007/s11408-020-00346-4).
- Mba, J.C., Pindza, E. and Koumba, U. (2018), "A differential evolution copula-based approach for a multi-period cryptocurrency portfolio optimization", *Financial Markets and Portfolio Management*, Vol. 32 No. 4, pp. 399-418, doi: [10.1007/s11408-018-0320-9](https://doi.org/10.1007/s11408-018-0320-9).
- McNeil, A.J. and Frey, R. (2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance*, Vol. 7 Nos 3-4, pp. 271-300, doi: [10.1016/s0927-5398\(00\)00012-8](https://doi.org/10.1016/s0927-5398(00)00012-8).
- Novy-Marx, R. and Velikov, M. (2016), "A taxonomy of anomalies and their trading costs", *The Review of Financial Studies*, Vol. 29 No. 1, pp. 104-147, doi: [10.1093/rfs/hhv063](https://doi.org/10.1093/rfs/hhv063).
- Paoletta, M.S., Polak, P. and Walker, P.S. (2025), "Risk parity portfolio optimization under heavy-tailed returns and dynamic correlations", *Journal of Time Series Analysis*, Vol. 46 No. 2, pp. 353-377, doi: [10.1111/jtsa.12792](https://doi.org/10.1111/jtsa.12792).
- Pickands, J. (1975), "Statistical inference using extreme order statistics", *The Annals of Statistics*, Vol. 3 No. 1, pp. 119-131, doi: [10.1214/aos/1176343003](https://doi.org/10.1214/aos/1176343003).
- Politis, D.N. and Romano, J.P. (1994), "The stationary bootstrap", *Journal of the American Statistical Association*, Vol. 89 No. 428, pp. 1303-1313, doi: [10.1080/01621459.1994.10476870](https://doi.org/10.1080/01621459.1994.10476870).
- Rigamonti, A. (2024), "Can machine learning make technical analysis work?", *Financial Markets and Portfolio Management*, Vol. 38 No. 3, pp. 399-412, doi: [10.1007/s11408-024-00451-8](https://doi.org/10.1007/s11408-024-00451-8).
- Rockafellar, R.T. and Uryasev, S. (2000), "Optimization of conditional value-at-risk", *Journal of Risk*, Vol. 2 No. 3, pp. 21-41, doi: [10.21314/jor.2000.038](https://doi.org/10.21314/jor.2000.038).
- Rousseeuw, P.J. (1987), "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65, doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Sahamkhadam, M. and Stephan, A. (2023), "Portfolio optimization based on forecasting models using vine copulas: an empirical assessment for global financial crises", *Journal of Forecasting*, Vol. 42 No. 8, pp. 2139-2166, doi: [10.1002/for.3009](https://doi.org/10.1002/for.3009).
- Sahamkhadam, M., Stephan, A. and Östermark, R. (2018), "Portfolio optimization based on GARCH-EVT-Copula forecasting models", *International Journal of Forecasting*, Vol. 34 No. 3, pp. 497-506, doi: [10.1016/j.ijforecast.2018.02.004](https://doi.org/10.1016/j.ijforecast.2018.02.004).
- Sarkar, P., Khanapuri, V.B. and Tiwari, M.K. (2025), "Integration of prediction and optimization for smart stock portfolio selection", *European Journal of Operational Research*, Vol. 321 No. 1, pp. 243-256, doi: [10.1016/j.ejor.2024.08.027](https://doi.org/10.1016/j.ejor.2024.08.027).
- Schubert, E. and Rousseeuw, P.J. (2019), "Faster k-medoids clustering: improving the PAM, CLARA, and CLARANS algorithms", *Similarity Search and Applications: 12th International Conference, SISAP 2019*, Cham, Springer, pp. 171-187. doi: [10.1007/978-3-030-32047-8_16](https://doi.org/10.1007/978-3-030-32047-8_16).
- Sheng, J., Chen, L., Chen, H. and An, Y. (2025), "CVaR-based risk parity model with machine learning", *Pacific-Basin Finance Journal*, Vol. 93, 102857, doi: [10.1016/j.pacfin.2025.102857](https://doi.org/10.1016/j.pacfin.2025.102857).
- Silva, N.F., Andrade, L.P., da Silva, W.S., de Melo, M.K. and Tonelli, A.O. (2024), "Portfolio optimization based on the pre-selection of stocks by the support vector machine model", *Finance Research Letters*, Vol. 61, 105014, doi: [10.1016/j.frl.2024.105014](https://doi.org/10.1016/j.frl.2024.105014).
- Simo-Kengne, B.D., Ababio, K.A., Mba, J. and Koumba, U. (2018), "Behavioral portfolio selection and optimization: an application to international stocks", *Financial Markets and Portfolio Management*, Vol. 32 No. 3, pp. 311-328, doi: [10.1007/s11408-018-0313-8](https://doi.org/10.1007/s11408-018-0313-8).
- Steinborn, M. and Liebscher, E. (2025), "Estimating portfolio risk with product copulas: a GARCH-EVT approach applied to financial data", *Journal of Economic Analysis*, Vol. 4 No. 4, pp. 14-33, doi: [10.58567/jea04040002](https://doi.org/10.58567/jea04040002).

-
- Tasche, D. (2002), "Expected shortfall and beyond", *Journal of Banking and Finance*, Vol. 26 No. 7, pp. 1519-1533, doi: [10.1016/S0378-4266\(02\)00272-8](https://doi.org/10.1016/S0378-4266(02)00272-8). *Journal of Capital Markets Studies*
- Wolff, D. and Echterling, F. (2024), "Stock picking with machine learning", *Journal of Forecasting*, Vol. 43 No. 1, pp. 81-102, doi: [10.1002/for.3021](https://doi.org/10.1002/for.3021).

Corresponding author

Juan C. Roca can be contacted at: jcroca@uhu.es
