A US Long-Short Equity Portfolio: Tail-Weighted Forecasts, Adverse-Tail Risk, and Sensitivity to Delisting Assumptions

QUARCC
Quantitative Research Note
November 2026

Abstract. We study a monthly, dollar-neutral US equity portfolio and a set of alternative specifications. Annual walk-forward Ridge and tail-weighted XGBoost models score 147 characteristics from Jensen, Kelly, and Pedersen (2023) across 529,082 stock-months. A convex program with an alpha-alignment factor, an adverse-tail penalty, and explicit trading and borrow costs forms the book. In an event-driven simulation from $10 million over January 2021 to August 2026 (68 months), the portfolio has a net information ratio of 3.77 against T-bill $+4\%$, 12.4% volatility, a -4.5% maximum drawdown, and $18.19\times$ growth. These figures should not be read as achievable returns. Models are walk-forward, but the design was chosen after 65 trials that had seen the evaluation window (deflated Sharpe ratio 0.886). About 91% of short gross sits in micro- and nano-capitalisation stocks, borrow availability is modelled and not observed, and liquidity costs grow quickly with assets. Forecast skill sits almost entirely on the short side and appears in the median stock, not the mean. The improvement from correcting the treatment of delisted stocks depends on the assumed delisting return. No alternative specification exceeds the baseline, and a preregistered synthetic test finds no forecast contribution in a joint risk system.

Data and Walk-Forward Protocol

The panel holds 529,082 stock-months of US common equities from January 2015 to August 2026, with 147 characteristics from Jensen, Kelly, and Pedersen (2023; hereafter JKP). Each month, with information dated at or before the previous month-end, models score the universe. They are refit every January on all earlier months. The evaluation window is the 68 months from January 2021 to August 2026. Each monthly portfolio is simulated in an event-driven engine (NautilusTrader) from $10 million with whole shares, brokerage fees, spread, impact, borrow, and recall.

Forecast and Alpha

Blend. Following Gu, Kelly, and Xiu (2020), the signal averages cross-sectional z-scores of a Ridge regression ($\alpha = 1000$) and a tail-weighted XGBoost model (400 trees, depth 6, Platt-calibrated, trained on names priced at $1 or more):

$$z_{i,t} = \mathrm{clip}\Big(z_{\mathrm{CS}}\big(\tfrac{1}{2} z_{\mathrm{CS}}(\hat{y}^{\mathrm{ridge}}_{i,t}) + \tfrac{1}{2} z_{\mathrm{CS}}(\hat{y}^{\mathrm{xgb}}_{i,t})\big), -3, 3\Big). \tag{1}$$

Candidates are the 200 highest and 200 lowest scores among names priced at $1 or more, for which a 1% position is at most 10% of average daily volume and 5% of market capitalisation.

Alpha. A conformalised quantile model (Romano, Patterson, and Candès, 2019) sets a confidence weight, and alpha follows the fundamental law (Grinold and Kahn, 2000):

$$\alpha_{i,t} = \mathrm{IC}^{+}_{\mathrm{leg}(i),t}\, \sigma_{i,t}\, z_{i,t}\, \mathrm{conf}_{i,t}, \tag{2}$$

where $\mathrm{IC}^{+}_{\mathrm{leg},t}$ is the leg's rank IC over the previous 36 months, floored at zero, and $\sigma_{i,t}$ is model volatility.

Asymmetric predictability. Within the 200-name short pool, rank IC is $+0.062$ in 2016–2020 and $+0.142$ in 2021–2026. Within the long pool it is -0.012 and -0.011. As a result $\mathrm{IC}^{+}_{\mathrm{long},t}$ is zero in 82% and 91% of months respectively, and longs mainly hedge the factor, beta, and industry exposure of the shorts. In the evaluation window, shorts contribute $+44.8\%$ a year and longs $+9.8\%$. With a negative long-pool IC, long P&L should not be read as evidence of selection skill.

Median and mean returns. The bottom-quintile blend score predicts the median stock, about 1.3–1.4% a month below the universe ($t = -3.5$), but the mean difference is zero ($t = 0.0$). Forecasts identify typical losers, and rare large winners in the same names offset them. Position caps and tail penalties address this right-tail exposure. Platt calibration also fits an inverted slope in 2020, which indicates instability in the XGBoost score.

Risk and Portfolio Construction

Covariance. A fundamental factor covariance $\mathbf{\Sigma}_{F,t}$ is augmented with an alpha-alignment factor (Ceria, Saxena, and Stubbs, 2012) along unit-norm alpha $\hat{\boldsymbol{\alpha}}_t$, rescaled by a variance correction $g_t \in [0.5, 4]$ learned from the portfolio's realised and predicted variance, and given a diagonal adverse-tail term:

$$\mathbf{\Sigma}_t = g_t\big(\mathbf{\Sigma}_{F,t} + \sigma^2_{\mathrm{AAF},t}\hat{\boldsymbol{\alpha}}_t\hat{\boldsymbol{\alpha}}_t^{\top}\big) + \mathrm{diag}(\mathbf{h}_{\mathrm{tail},t}). \tag{3}$$

Here $h_{\mathrm{tail},i,t}$ is built from the 0.99 conformal quantile of losses on the held side, namely squeezes for shorts and crashes for longs, in excess of the Gaussian quantile. It is scaled by a power-law conditional tail-variance multiplier (Taleb, 2020).

Costs. Half-spreads scale the high–low range (Corwin and Schultz, 2012) by a factor of 0.3947, calibrated on the most liquid names in 2016–2020. Impact follows Almgren et al. (2005), with terms in $|\Delta w_i|^{1.6}$ and Δw_i^2 evaluated at current NAV. Shorts pay a borrow fee tiered by dollar volume (D'Avolio, 2002) plus expected recall cost.

Program. Each month the portfolio solves

$$\max_{\mathbf{w}} \; \boldsymbol{\alpha}_t^\top \mathbf{w} - \lambda m_t \mathbf{w}^\top \boldsymbol{\Sigma}_t \mathbf{w} - \mathrm{TC}_t(\Delta \mathbf{w}) - \mathrm{Carry}_t(\mathbf{w}^-) - \tau \|\mathbf{w} - \mathbf{w}_{\mathrm{drift}}\|_1, \tag{4}$$

with $\lambda = 2$, drawdown multiplier $m_t = 1 + 40\,\mathrm{DD}_{t,6\mathrm{m}}$, and $\tau = 0.01$. Long and short legs are each exactly 50% of NAV. Beta on Fama–MacBeth up and down market betas is within ± 0.05, the portfolio may not be short market convexity (Ang, Chen, and Xing, 2006), Fama–French 49-industry net exposure is within $\pm 5\%$, and single names are capped at 2.5%. We solve with Clarabel (Goulart and Chen, 2024) through CVXPY (Diamond and Boyd, 2016).

Results

Table 1 reports the simulated results. Costs are about 5.2% a year: 1.8% trading (commission, spread, and impact) and 3.4% borrow and recall. The portfolio holds 410 names on average (263 long, 146 short) and exceeds 500 names in 4 of 68 months because of small residual positions (maximum 509). The ten largest winners account for 22% of gross profit and loss across 1,847 names traded.

Table 1. Results, January 2021 to August 2026 (68 months), net of modelled costs, simulated from initial capital of $10 million. Sharpe and alpha use monthly excess returns; alpha's t-statistic uses Newey–West errors. S&P 500 figures use the price index.

Metric	Portfolio	S&P 500
Growth of $1	18.19	2.05
CAGR / arithmetic mean	66.8% / 53.0%	13.5% / 13.8%
Volatility	12.36%	15.10%
Sharpe ratio	4.10	0.70
Maximum drawdown	−4.49%	−24.8%
IR vs. T-bill +4%	3.77	0.44
Beta / alpha to S&P 500 (t)	+0.09 / 48.9% (8.35)	—
Rolling 12-month beta	−0.31 to +0.99	—
Costs per year	5.2%	—
Monthly one-way turnover	13.9%	—

Year	Portfolio	S&P 500
2021	17.3%	26.9%
2022	65.2%	−19.4%
2023	76.4%	24.2%
2024	97.7%	23.3%
2025	106.5%	16.4%
2026 (to Aug.)	30.3%	12.3%

Table 2. Sensitivity of results to universe construction, exit liquidity, and missing-return treatment. Event-driven simulations from $10 million, January 2021 to August 2026. All rows share parameters and rebalance rules. Maximum drawdown is on excess returns. The bold row is the baseline specification. The universe labelled "all rows" scores every stock at formation, including those whose next-month return is missing.

Universe, exit liquidity	Missing return treatment	IR	Growth	Max DD	Beta	Costs/yr
Original, original exits	−30% long, 0 short	3.631	17.81	−5.1%	+0.118	5.13%
Original, repaired exits	−30% long, 0 short	3.599	17.95	−5.2%	+0.123	5.06%
Original, repaired exits	Last price	3.613	18.01	−5.2%	+0.123	5.06%
All rows, original exits (not kept)	−30% long, 0 short	3.444	16.27	−4.7%	+0.114	6.91%
All rows, repaired exits	−30% long, 0 short	3.602	17.52	−4.7%	+0.109	5.11%
All rows, repaired exits	**Last price**	**3.769**	**18.19**	**−4.6%**	**+0.093**	**5.18%**

Sensitivity to Universe and Delisting Treatment

Three data-handling choices materially affect the results (Table 2).

Look-ahead in the universe. Scoring only stocks with an observed next-month return conditions on information unavailable at formation and excludes many stocks about to delist. Scoring all stocks restores 2,311 name-months (482,241 to 484,552) and leaves trained models unchanged.

Exit liquidity. Assigning a delisted name the smallest volume among requested names can make an exit cost up to 244% of notional. In the original universe this error is mostly hidden, whereas scoring all stocks exposes it (information ratio 3.44). The baseline uses the last observed spread and liquidity.

Missing returns. One treatment books a held stock with no return at −30% when long and 0 when short. The baseline closes both at the last available price, a return of zero. This affects 189 positions: 147 long and 42 short. Shumway (1997) finds that delisting returns are materially negative for performance-related delistings, while others, such as cash mergers, close near the last price, so the true effect lies between the two treatments. The entire improvement over the original specification comes from this assumption: at −30% the corrected information ratio is 3.60, below the original 3.63. The −30% charge is worth 1.99% a year. We therefore regard 3.60–3.77 as the plausible range.

Limitations

Selection. Models are walk-forward, and the turnover penalty was selected on 2018–2020. However, the design passed through 65 trials in which 2021–2026 results for earlier versions were visible, which gives a deflated Sharpe ratio (Bailey and López de Prado, 2014) of 0.886, below the conventional 0.95. The data-handling corrections were also

made with results in view. The sample is therefore not a clean holdout.

Short-book composition. By gross, 91% of the short side is in micro- and nano-capitalisation stocks (54% nano), 61% trades below \$5, and the median short has a \$58m market capitalisation. By the model's dollar-volume tiers, 34% is hard to borrow or lower; the lowest tier is excluded only at entry, not while held. The model charges 3.4% a year for borrow and recall but does not observe loan availability or fee history, which is the largest source of uncertainty for implementation.

Capacity. All results assume \$10 million of capital and do not scale linearly with assets. On the development window, a closely related 500-name control has 34–42% of short gross beyond one day of 10% of daily volume and 13–18% beyond five days at that capital. Costs grow faster than gross, and net excess return stops rising at about 100–120% gross. We do not estimate returns at larger capital, but they would be materially lower.

Time-varying market exposure. Full-sample beta is $+0.09$ with a standard error near 0.09, but rolling 12-month beta ranges from -0.31 to $+0.99$.

Simulated execution. All results are simulated with modelled costs, and no order was executed.

Alternative Specifications

Table 3 reports alternative specifications. No alternative exceeds the baseline information ratio in the evaluation window, and the three development-window specifications in the table fall below the control. These negative results indicate where the predictability does and does not lie.

Table 3. Alternative specifications. "Evaluation" is January 2021 to August 2026, the same window as Table 1; "Development" is July 2016 to December 2020. Dashes denote quantities not measured.

Variant	Window	IR	Vol.	Max DD	Beta	Outcome
500-name control	Evaluation	3.478	12.4%	−4.6%	+0.098	500-name limit; excess 47.1% a year
Normal-rank score blend	Evaluation	3.555	—	—	—	−2.12% a year vs. control ($t = -1.01$)
Volatility target 8%	Evaluation	3.697	8.7%	−3.5%	+0.065	−11.1% a year vs. control ($t = -3.58$)
Separated μ, Σ, cost	Evaluation	2.29	29.8%	−28.8%	+0.77	48.7× growth; high beta and drawdown
Unified compiler (best)	Development	1.04	7.3%	−6.0%	—	Below control's 1.36
Fractional IR maximiser	Development	0.24	12.7%	−12.4%	—	$z = -2.26$ vs. control
LightGBM mean forecast	Development	0.62	31.6%	−57%	+0.33	Below control's 1.36
Direct parametric policy	Evaluation	—	0.44%	−1.1%	—	Near cash; Sharpe 0.29, gross 9–18%

Volatility targeting. Scaling to an 8% target (Moreira and Muir, 2017) lowers risk but lowers return by more. On the development window, at matched risk, the target arm and the control lie on the same frontier: Sharpe ratio peaks near 8% volatility for both (1.94 and 1.91), so the higher Sharpe ratio of the arm over the full-size control (1.77) reflects smaller scale.

Forecast construction. A normal-rank blend removes the unstable calibration but trails the control by a statistically insignificant 2.12% a year. A separated forecast, risk, and cost system compounds faster (48.7×), with volatility of 29.8% and a beta of $+0.77$.

Integrated optimisation. Unified compilers (Boyd et al., 2024), a fractional information-ratio maximiser (Dinkelbach, 1967), and a direct policy (Brandt, Santa-Clara, and Valkanov, 2009) all fall below the control. Policies linear in rank-normalised characteristics spread their bets, whereas the predictability here is concentrated.

Placebo test of a joint risk system. We also test a joint risk system with cell-calibrated forecasts in a synthetic economy with planted squeezes and delistings. On a real development window (July 2016 to December 2020), size-quintile specific-risk shrinkage gains $+0.92\%$ a month (95% CI $[-0.62\%, +2.45\%]$, $p = 0.229$; stepdown-adjusted $p = 0.102$; Romano and Wolf, 2005), with the same Sharpe ratio and more risk. In a preregistered synthetic test over 144 months against 16 shuffled-forecast placebos with identical risk, cost, and constraints, the mean difference is -0.865% a month (95% CI $[-2.43\%, +0.70\%]$, $p = 0.283$), the system ranks 16th of 16 (Monte Carlo $p = 0.941$), and an anytime-valid e-process (Ramdas et al., 2023) ends at $\log E = -0.46$. In this setting, positive returns arise from the risk, cost, and constraint structure and not from forecast ordering. The system was not evaluated on 2021–2026 data.

Code and Data Availability

The research code is not publicly available, and the data are licensed and cannot be redistributed.

References

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li, 2005, Direct estimation of equity market impact, *Risk* 18, 58–62.

Ang, Andrew, Joseph Chen, and Yuhang Xing, 2006, Downside risk, *Review of Financial Studies* 19, 1191–1239.

Bailey, David H., and Marcos López de Prado, 2014, The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality, *Journal of Portfolio Management* 40, 94–107.

Boyd, Stephen, Kasper Johansson, Ronald Kahn, Philipp Schiele, and Thomas Schmelzer, 2024, Markowitz portfolio construction at seventy, *Journal of Portfolio Management* 50, 117–160.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *Review of Financial Studies* 22, 3411–3447.

Ceria, Sebastián, Anureet Saxena, and Robert A. Stubbs, 2012, Factor alignment problems and quantitative portfolio management, *Journal of Portfolio Management* 38, 29–43.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–760.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Diamond, Steven, and Stephen Boyd, 2016, CVXPY: A Python-embedded modeling language for convex optimization, *Journal of Machine Learning Research* 17, 1–5.

Dinkelbach, Werner, 1967, On nonlinear fractional programming, *Management Science* 13, 492–498.

Goulart, Paul J., and Yuwen Chen, 2024, Clarabel: An interior-point solver for conic programs with quadratic objectives, arXiv preprint arXiv:2405.12762.

Grinold, Richard C., and Ronald N. Kahn, 2000, *Active Portfolio Management*, 2nd ed., McGraw-Hill, New York.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Jensen, Theis I., Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Ramdas, Aaditya, Peter Grünwald, Vladimir Vovk, and Glenn Shafer, 2023, Game-theoretic statistics and safe anytime-valid inference, *Statistical Science* 38, 576–601.

Romano, Joseph P., and Michael Wolf, 2005, Stepwise multiple testing as formalized data snooping, *Econometrica* 73, 1237–1282.

Romano, Yaniv, Evan Patterson, and Emmanuel J. Candès, 2019, Conformalized quantile regression, *Advances in Neural Information Processing Systems* 32, 3543–3553.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Taleb, Nassim Nicholas, 2020, *Statistical Consequences of Fat Tails*, STEM Academic Press.