QUARCC Purple: Encompassing LASSO, Semivariance Risk, and Relax-and-Round Construction in a Market-Neutral US Equity Fund

QUARCC Purple team
McGill–FIAM Asset Management Hackathon 2026, Final Round
November 2026

Abstract. We describe QUARCC Purple, a simulated $1 billion US equity fund built for the final round of the McGill–FIAM Asset Management Hackathon 2026 under a market-neutral mandate benchmarked to the 3-month T-bill plus 4% a year. Monthly walk-forward forecasts come from a value-weighted Encompassing LASSO (Han, He, Rapach, and Zhou, 2019) over 147 characteristics from Jensen, Kelly, and Pedersen (2023), augmented with a Stambaugh, Yu, and Yuan (2012) mispricing composite, a machine-learning forecast of next-quarter earnings surprises, and publication-lagged short interest. Targets are CAPM residual ranks. A horizon-matched factor model with directional specific semivariances prices risk, and a relax-and-round quadratic program forms a book twice a month that is traded over five sessions. From January 2021 to August 2026 (68 months, 1,421 sessions), after modelled costs, the fund grows $1.000bn to $1.467bn: a 7.00% CAGR, 3.96% volatility, Sharpe ratio 0.91, maximum drawdown -2.99%, and S&P 500 beta 0.012. It does *not* beat its benchmark: the information ratio against T-bill $+4\%$ is -0.11 (standard error 0.42). The forecasts are walk-forward, but construction settings were chosen with the evaluation window visible, so these results are a simulation rather than a clean holdout or a live record.

Mandate, Data, and Point-in-Time Clocks

The mandate requires 200–500 names, gross exposure at most 200%, net exposure within $\pm 50\%$, a 5% single-name limit, sector net exposure within $\pm 20\%$, and no short position below a $500m market capitalisation. The benchmark is the 3-month T-bill plus 4% a year, compounded monthly.

The sample consists of US equities with the 147 characteristics of Jensen, Kelly, and Pedersen (2023; hereafter JKP) and daily stock returns. Every input enters the model at the time it became public: SEC filings at their acceptance time, rolled to the next session when accepted after the close, and exchange short interest eight trading days after settlement. Observations flagged as revised are excluded because their first-reported values are unavailable, and forward-looking split factors are not used.

At each month-end the estimation universe requires $|\mathrm{prc}_{i,t}| \geq \$5$ and 126-day average daily dollar volume of at least $2m. Short eligibility additionally requires market equity of at least $500m, and candidate positions require at least $3bn. A book is formed at each month-end and at the first session on or after the 15th of the following month, giving 136 rebalances. Trading starts one session after formation and runs over five sessions.

Signals and Encompassing LASSO Forecaster

Three added signals ($147 \to 150$ predictors).

1. *Mispricing composite.* Following Stambaugh, Yu, and Yuan (2012), we average the signed uniform ranks of ten anomalies available in the panel, omitting distress: O-score, share issuance, net equity issuance, operating accruals, net operating assets, asset growth, and investment (negative sign); and momentum, gross profitability, and return on assets (positive sign). A stock needs at least five of the ten.
2. *Earnings-surprise forecast.* Bernard and Thomas (1989) condition on realised surprises. We instead train an annual walk-forward LightGBM regressor on the 147 characteristics to predict next quarter's as-filed standardised unexpected earnings before the announcement.
3. *Short interest.* Following Boehmer, Huszar, and Jordan (2010), we use short interest over shares outstanding at the latest settlement date that is public by the formation close.

Standardisation and mid-month refresh. Following Green, Hand, and Zhang (2017), each predictor is winsorised at its 1st and 99th cross-sectional percentiles and standardised, with missing values set to zero:

$$z_{i,j,t} = \frac{\mathrm{clip}\big(x_{i,j,t},\, Q_{0.01}(x_{\cdot,j,t}),\, Q_{0.99}(x_{\cdot,j,t})\big) - \mu_{j,t}}{\sigma_{j,t}}. \tag{1}$$

At the mid-month formation T_{mid} we recompute nine 21-day price signals (one-month return, volatility, idiosyncratic volatility, maximum daily returns, and skewness measures) from daily returns and substitute them, holding month-end coefficients fixed.

Residual rank target. Gu, Kelly, and Xiu (2020) predict excess returns, which load on high-beta tails. Extending the residual-return principle of Blitz, Huij, and Martens (2011), the label is the centred cross-sectional percentile rank of the CAPM beta-adjusted return:

$$\begin{aligned} \tilde{r}_{i,t+1} &= r^{e}_{i,t+1} - \beta^{60\mathrm{m}}_{i,t}\, r^{e}_{m,t+1}, \\ y_{i,t+1} &= 100\,\big(\mathrm{PctRank}_{t}(\tilde{r}_{i,t+1}) - \tfrac{1}{2}\big)\,. \end{aligned} \tag{2}$$

Value-weighted Encompassing LASSO. Building on Rapach, Strauss, and Zhou (2010) and Han et al. (2019):

1. *Weights and penalty selection.* Each monthly regression weights stock i in proportion to market equity, rescaled to mean one, so microcaps cannot dominate. Coordinate descent (Friedman, Hastie, and Tibshirani, 2010) traces a λ path and selects $\lambda^{\star}_{\tau}$ by the corrected AIC of Hurvich and Tsai (1989).

2. *Two branches.* S-LASSO fits all predictors jointly and averages coefficients over the expanding history, $\bar{\boldsymbol{\beta}}_t^{\mathrm{S}} = \frac{1}{t-1}\sum_{\tau<t}\hat{\boldsymbol{\beta}}_\tau^{\mathrm{S}}$. C-LASSO forms out-of-sample univariate forecasts $u_{i,j,\tau}$, selects the set $\mathcal{S}_\tau$ with a non-negative LASSO on those forecasts, and equal-weights the survivors: $\hat{y}_{i,t}^{\mathrm{C}} = |\mathcal{S}_{t-1}|^{-1}\sum_{j\in\mathcal{S}_{t-1}} u_{i,j,t}$.
3. *Forecast encompassing.* Following Chong and Hendry (1986), let $e_{i,\tau}^{\mathrm{S}}$ be the S-LASSO error and $d_{i,\tau} = \hat{y}_{i,\tau}^{\mathrm{C}} - \hat{y}_{i,\tau}^{\mathrm{S}}$. With value-weighted demeaned series $\tilde{d}$ and $\tilde{e}^{\mathrm{S}}$,

$$\begin{aligned}\hat{\lambda}_\tau &= \operatorname{clip}\!\left(\frac{\sum_i v_{i,\tau}\tilde{d}_{i,\tau}\tilde{e}_{i,\tau}^{\mathrm{S}}}{\sum_i v_{i,\tau}\tilde{d}_{i,\tau}^2},\,0,\,1\right),\\ \hat{y}_{i,t} &= (1-\bar{\lambda}_t)\,\hat{y}_{i,t}^{\mathrm{S}} + \bar{\lambda}_t\,\hat{y}_{i,t}^{\mathrm{C}},\end{aligned} \tag{3}$$

   where $\bar{\lambda}_t$ is the expanding mean of $\hat{\lambda}_\tau$.

Market-state dial and horizon scaling. Extending Pollet and Wilson (2010), each formation falls in one cell of a 3×2 grid: terciles of the 63-day correlation share (equal-weighted index variance over mean stock variance, with cuts from past data only) crossed with the sign of the trailing 12-month HML return. Using credibility shrinkage in the spirit of Bühlmann and Straub (1970) with prior weight $N_0 = 12$, the cell's past rank IC is shrunk toward the unconditional rank IC:

$$\rho_t = \operatorname{clip}\!\left(\frac{n_{\mathrm{cell}}\,\overline{\mathrm{IC}}_{\mathrm{cell},t} + N_0\,\overline{\mathrm{IC}}_{\mathrm{all},t}}{(n_{\mathrm{cell}}+N_0)\,\overline{\mathrm{IC}}_{\mathrm{all},t}},\,0,\,1.5\right). \tag{4}$$

The dial is 1 until 36 past formations exist or when the unconditional IC is at most 0.005. With median-centred score $\tilde{m}_{i,t}$, scale $s_f = 0.0484$, and decay $\kappa = 0.00217$, in the spirit of Gârleanu and Pedersen (2013), the H-session forecast is

$$\mu_{i,t}^{(H)} = \rho_t\, s_f\, \frac{\tilde{m}_{i,t}}{100}\,\frac{1-e^{-\kappa H}}{1-e^{-21\kappa}}. \tag{5}$$

Factor and Semivariance Risk Model

At each formation the H-session covariance is $\boldsymbol{\Sigma}_t^{(H)} = \mathbf{B}_t\mathbf{F}_t^{(H)}\mathbf{B}_t^\top + \mathbf{D}_t^{(H)}$, built in six steps.

1. *Hedge beta.* Whereas Vasicek (1973) shrinks toward one cross-sectional mean, we shrink 252-day betas toward their GICS sector mean and average the result equally with the 60-month beta.
2. *Factor returns.* Daily returns are regressed on one beta column, GICS sector indicators, and five rank-standardised styles (size, value, momentum, profitability, investment) by $\sqrt{\mathrm{ME}}$-weighted least squares, with a cap-weighted sector sum-to-zero constraint (Menchero, Orr, and Wang, 2011).
3. *Horizon covariance.* With an 84-day half-life, lag-j autocovariances $\boldsymbol{\Gamma}_j$ extend Newey and West (1987) to the holding horizon through

$$\widetilde{\mathbf{F}}^{(H)} = H\boldsymbol{\Gamma}_0 + \sum_{j=1}^{H-1}(H-j)\big(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j^\top\big), \tag{6}$$

   followed by an eigenvalue floor at $10^{-8}\lambda_{\max}$ and 20% diagonal shrinkage.
4. *Specific risk.* On 2017–2020 formations a Patton (2011) QLIKE race with Diebold and Mariano (1995) statistics selects between HAR-RV (Corsi, 2009) and EWMA forecasts (half-lives 84 and 168 days), with 25% shrinkage toward the sector median and a floor at the 5th percentile. We also estimate separate downside and upside residual semivariances, $d_{i,t}^-$ and $d_{i,t}^+$.
5. *Volatility regime and residual factors.* A 42-day half-life volatility regime adjustment rescales factor and specific variances. Following Connor and Korajczyk (1986), five principal components of 252-day residuals are appended to $\mathbf{B}_t$.
6. *Stress mixture.* A two-state Markov chain in the spirit of Hamilton (1989), on 21-day realised market variance, gives a stress probability p_t and $\mathbf{F}_t \leftarrow (1-p_t)\mathbf{F}_t + p_t\mathbf{F}_{\mathrm{stress},t}$. The stressed matrix moves correlations halfway to ± 1 and scales factor volatilities by 1.5.

Portfolio Construction and Execution

Signed QP. With $\mathbf{q} = \mathbf{w} - \mathbf{w}_{\mathrm{pre}}$, the final stage solves

$$\begin{aligned}\min_{\mathbf{w}}\;& -\boldsymbol{\mu}_t^{(H)\top}\mathbf{w} + \frac{\gamma}{2}\Big(\mathbf{w}^\top\mathbf{B}_t\mathbf{F}_t^{(H)}\mathbf{B}_t^\top\mathbf{w} + H\sum_i d_{i,t}^{\pm}w_i^2\Big)\\ &+ \lambda_\beta\big(\boldsymbol{\beta}_{\mathrm{hedge}}^\top\mathbf{w}\big)^2 + \frac{H}{252}\mathbf{b}_t^\top\mathbf{w}^- + \sum_i\frac{\lambda_u H d_{i,t}}{c_{i,t}}w_i^2\\ &+ \mathrm{TC}(\mathbf{q}) + M_1\|\boldsymbol{\xi}_1\|_1 + M_2\|\boldsymbol{\xi}_2\|_1,\end{aligned} \tag{7}$$

with $\gamma = 0.25$, $\lambda_\beta = 10$, feature-coverage penalty $\lambda_u = 0.25$ ($c_{i,t}\in[0.1,1]$ is the share of non-missing characteristics), and exact slack penalties $M_1 = 5$ and $M_2 = 1$. Here $d_{i,t}^{\pm}$ is the downside semivariance for longs and the upside semivariance for shorts, floored at half the symmetric variance. An earlier free-sign stage uses $\gamma = 1$. To handle the 500-name limit and a 10 bp minimum position, we solve the continuous relaxation over eligible candidates, keep at most 440 names, lock their signs, and re-solve. Sixty names of headroom are reserved for the overlap of exits and entries. All QPs are solved with Clarabel (Goulart and Chen, 2024) through CVXPY (Diamond and Boyd, 2016).

Constraints. The core problem imposes gross exposure at most 180%; sector net within $\pm 1\%$; style bands of $0.01/\sigma_{\mathrm{factor},k}$; ex-ante volatility at most 12% and stressed volatility at most 24%; long caps of 4.25% and short caps of 1%, each also limited by five days at 10% of average daily volume; and beta limits on three estimators: $|\boldsymbol{\beta}_{\mathrm{hedge}}^\top\mathbf{w}| \le 0.02$, $|\boldsymbol{\beta}_{60\mathrm{m}}^\top\mathbf{w}| \le 0.05$, and $|\boldsymbol{\beta}_{\mathrm{WD}}^\top\mathbf{w}| \le 0.05$, where $\boldsymbol{\beta}_{\mathrm{WD}}$ averages the Welch (2022) and Dimson (1979) betas.

Costs and borrow. Half-spreads scale each stock's 21-day high–low estimate (Corwin and Schultz, 2012) by a ratio calibrated on 2015–2020 TAQ effective spreads within size

groups. Executions pay brokerage and regulatory fees, the half-spread, and square-root impact (Almgren et al., 2005), which the QP approximates with a matched quadratic term. Short borrow is tiered on short interest over shares (D'Avolio, 2002): 30 bps a year below 5%, then 1%, 3%, and 8% at the 5%, 10%, and 20% thresholds.

Final portfolio and trade schedule. The final portfolio scales core targets by 1.08, with formation gross capped at 188% and caps of 3.9% long and 1.08% short. Trades run in equal fifths over five sessions, with a 10% daily ADV limit. Guards trim gross from 190% to 185% and any long from 4.4% to 3.9%, and cover any short whose capitalisation falls below a $550m buffer. Eight local language-model agents read filings, draft commentary, measure crowding, and critique the portfolio; none of them determines a portfolio weight.

Table 1. Headline results, 4 January 2021 to 31 August 2026 (68 months, 1,421 sessions), after modelled costs. Volatility, Sharpe, Sortino, beta, and alpha use monthly returns; Sortino uses a 0% target. S&P 500 figures use the price index.

Metric	QUARCC Purple	T-bill +4%	S&P 500
Ending value of $1.000bn	$1.467bn	$1.509bn	$2.046bn
CAGR	7.00%	7.53%	13.47%
Volatility	3.96%	—	15.10%
Sharpe ratio	0.91	—	0.70
Sortino ratio	4.21	—	1.51
Maximum drawdown	−2.99%	—	−25.43%
Ulcer index	0.52%	—	8.10%
Beta to the S&P 500 (s.e.)	0.012 (0.030)	—	1.00
Alpha to the S&P 500 (t)	3.47% a year (2.45)	—	—
Information ratio vs. T-bill +4% (s.e.)	−0.11 (0.42)	—	—
Worst month / worst day	−1.43% / −1.37%	—	—
Months positive	48 of 68	—	—

Year	Fund	T-bill +4%
2021	6.98%	4.12%
2022	1.34%	6.19%
2023	15.10%	9.45%
2024	11.16%	9.34%
2025	2.08%	8.37%
2026 (to Aug.)	3.62%	5.21%

Results

Table 1 reports the results. No mandate limit is breached on any day of the sample. The fund holds 246–357 names, with gross exposure averaging 136.5% (maximum 192.9%, final 186.0%). Single-name exposure reaches at most 4.6%, and net exposure stays between −3.1% and +2.2%.

Net profit of $467.07m comprises $257.7m from stock positions and $239.5m of interest on cash, less $30.1m of costs: $9.5m of trading (14 bps a year) and $20.6m of borrow (29 bps a year). Interest is nearly half of gross profit, so a T-bill-based benchmark is the appropriate comparison. In 2022, when the S&P 500 total return is −18.18%, the fund returns +1.34%.

Limitations

Benchmark shortfall. The 7.00% CAGR trails the 7.53% benchmark by 53 bps a year. Costs equal 72% of the $41.6m cumulative shortfall, and the information ratio before costs is approximately zero. With a standard error of 0.42, the data cannot distinguish a modest edge from no edge.

Selection. Forecasts are walk-forward, but construction settings, including the final gross scale, were chosen with 2021–2026 visible. The sample is therefore not a clean holdout, and Sharpe ratios and drawdowns should be read with this selection in mind.

Time variation. The fund trails the benchmark in 2022, 2025, and 2026 to date; in 2025 it returns 2.08% against 8.37%. Average gross exposure of 136.5% is well below the 200% limit, which is consistent with the optimiser charging entry costs against a horizon-scaled forecast.

Simulated execution. All results are simulated and no order was executed. Fills, borrow availability, and market impact at $1 billion may differ from the cost model.

Code and Data Availability

Code is available under the MIT License at github.com/NicholasXydis/QuarccPurple. The characteristic and return 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.

Bernard, Victor L., and Jacob K. Thomas, 1989, Post-earnings-announcement drift: Delayed price response or risk premium?, *Journal of Accounting Research* 27, 1–36.

Blitz, David, Joop Huij, and Martin Martens, 2011, Residual momentum, *Journal of Empirical Finance* 18, 506–521.

Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan, 2010, The good news in short interest, *Journal of Financial Economics* 96, 80–97.

Bühlmann, Hans, and Erwin Straub, 1970, Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 70, 111–133.

Chong, Yock Y., and David F. Hendry, 1986, Econometric evaluation of linear macro-economic models, *Review of Economic Studies* 53, 671–690.

Connor, Gregory, and Robert A. Korajczyk, 1986, Performance measurement with the arbitrage pricing theory, *Journal of Financial Economics* 15, 373–394.

Corsi, Fulvio, 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

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.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani, 2010, Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* 33, 1–22.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

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

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

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

Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.

Han, Yufeng, Ai He, David E. Rapach, and Guofu Zhou, 2019, Firm characteristics and expected stock returns, Working paper.

Hurvich, Clifford M., and Chih-Ling Tsai, 1989, Regression and time series model selection in small samples, *Biometrika* 76, 297–307.

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

Menchero, Jose, D. J. Orr, and Jun Wang, 2011, The Barra US Equity Model (USE4) methodology notes, MSCI Barra.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Patton, Andrew J., 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* 160, 246–256.

Pollet, Joshua M., and Mungo Wilson, 2010, Average correlation and stock market returns, *Journal of Financial Economics* 96, 364–380.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Vasicek, Oldrich A., 1973, A note on using cross-sectional information in Bayesian estimation of security betas, *Journal of Finance* 28, 1233–1239.

Welch, Ivo, 2022, Simply better market betas, *Critical Finance Review* 11, 37–64.