

Improving DCC–GARCH Covariance Estimator

Diagnosis, Correction, and Portfolio Implications

Xantos Labs LLC

September 2026

Abstract

The dynamic conditional correlation (DCC) covariance forecast used by the Xantos allocator and backtester was produced by `mgarch` version 0.3.0, the release pinned in both repositories’ `requirements.txt` and the latest published on PyPI. That release departs from the Engle–Sheppard specification in four ways, the most serious of which computes the correlation-targeting matrix $\bar{Q}$ from *raw returns reshaped rather than transposed* into asset-major order. The reshape scrambles the array, so the mean-reversion target of the correlation recursion is noise. Measured on data with a known correlation of 0.70, the release returns 0.10. Because the allocator minimises $w'Hw$, and that objective is driven almost entirely by off-diagonal terms, the covariance matrix the optimiser was handed carried little usable information about co-movement.

Version matters here. The project’s GitHub master branch computes the same quantity as `np.cov(self.rt.T)`, a genuine transpose, and does not exhibit this defect; master is ahead of the last PyPI release, so the published source and the installed code disagree. Every claim in this note about the reshape is a claim about 0.3.0, which is the code that ran in production. The remaining three defects are present in both.

A separate defect truncated the estimation window: a single `dropna(how='any')` across the universe let one short-history symbol collapse the sample for every other asset, in the live universe from 1,829 observations to 567.

This note derives the correct specification, documents both defects with reproducible evidence, describes the replacement estimator, and reports the effect on allocations and on a replication of the August 2022 study that first adopted DCC.

It then records the search for a change that improves crisis behaviour as well as return and Sharpe, which ended in a state-dependent no-trade band, and a fix for the allocation’s dependence on names it does not hold. Every figure is recomputed from the backtest outputs; interactive reports are linked in Section 12.

Contents

1	The DCC specification	4
2	The defects	5
2.1	D1: $\bar{Q}$ is computed on scrambled raw returns (0.3.0 only)	5
2.2	D2: (α, β) are never estimated (both versions)	6
2.3	D3: the horizon scales the covariance by $\sqrt{h}$ (both versions)	6
2.4	D4: no multi-step forecast (both versions)	6
2.5	D5: one short-history symbol truncates the whole window	6
3	The correction	6
3.1	A specification-faithful estimator	6
3.2	Per-pair estimation windows	7

4	Evidence	7
4.1	Correlation recovery	7
4.2	Horizon scaling	7
4.3	Implied volatilities	7
5	Implications	8
6	Backtest replication	8
6.1	Design	8
6.2	Configuration required to obtain valid results	9
6.3	The 2022 result replicates under the shipped estimator	9
6.4	The advantage does not survive correcting the estimator	9
6.5	What the shipped estimator was actually doing	9
6.6	Crisis behaviour	10
6.7	What survives	10
6.8	Caveats	10
6.9	Full results	11
6.10	Annualised net return	11
6.11	Annualised net volatility	11
6.12	Sharpe ratio	12
6.13	Maximum drawdown	12
7	A defect in the backtester’s rebalance trigger	12
7.1	What it was	12
7.2	Evidence	12
7.3	The fix	13
7.4	Impact	13
7.5	Scope and what it does not tell us	13
8	Crisis behaviour: searching for a DCC change that works	14
8.1	Goal and the gap	14
8.2	Thesis 1: the covariance forecast is the bottleneck	14
8.3	Thesis 2: the book trades too much	15
8.4	Refined thesis: it is when the book trades	15
8.5	The state-dependent band	16
8.6	Result	16
8.7	Testing the thesis directly	17
8.8	Why the state band was chosen, and what it costs	18
8.9	Decision	19
8.10	Caveats	19
9	Universe dependence	19
9.1	The observation	19
9.2	Diagnosis from first principles	19
9.3	Options considered	20
9.4	The fix: fit the dynamics on the top ten	20
9.5	Evidence	20
9.6	Backtest	21

10 Reducing the maximum drawdown	22
10.1 Where the drawdown comes from	22
10.2 What was tried	23
10.3 The change: a volatility-scaled trailing stop	23
10.4 Result	24
10.5 The controls are not specific to DCC	24
10.6 Re-entry variants that were rejected	25
10.7 Caveats	25
11 The live universe, December 2025 to August 2026	25
11.1 A spread has a feasibility ceiling	25
11.2 Result	26
12 What the estimator cannot fix	26
12.1 Thesis	27
12.2 The floor is unreachable a third of the time	27
12.3 When the floor is unreachable the book is a corner	28
12.4 Spread and cap are the risk dial	28
12.5 What this implies for the live tiers	29
13 A screen of allocator changes	30
13.1 Method	30
13.2 Theses and verdicts	30
13.3 Results	30
13.4 Following the two algorithms' own intent	31
13.5 What was falsified, and why	31
14 Two levers that the backtest rejected	32
14.1 Charging the objective for turnover	32
14.2 Asking for less return when the tape is stressed	33
14.3 What the backtest said	34
14.4 Why the screen was wrong	34
14.5 What this says about the harness	34
15 Where the return comes from	35
15.1 The return is five names	35
15.2 Both caps bind, nearly always	35
15.3 The after-tax lens, which corrects Section 10	36
15.4 What the tax gap is, and what it is not	36
15.5 Relaxing the group cap	36
15.6 Is that conviction, or three lucky names?	37
15.7 What was rejected	37
15.8 Does it rest on one name?	37
15.9 Recommendation, and what it is not	38
16 Interactive reports	38
17 Conclusions	39
18 Reproducing this	41

Summary: theses, conclusions and decisions

This note records three pieces of work. Each was run as a thesis stated before the test.

- **The estimator (Sections 1 to 6).** Thesis: the DCC covariance used since 2022 is not the Engle–Sheppard model. Confirmed: four defects in mgarch 0.3.0 plus a window truncation. Corrected, DCC’s edge over the GCLPM baseline at a 30% spread falls from +3.48pp to −1.22pp on the re-based runs. The 2022 adoption result was produced by the defect.
- **A defect in the backtester itself (Section 7).** Its rebalance trigger compared two weight vectors by the sum of their difference, which is always zero, so roughly a third of large target changes were never traded – in the backtest only; live trading was unaffected. Every 30% figure here is re-based on the fixed engine.
- **Crisis behaviour (Section 8).** Goal: beat the baseline on return, Sharpe and crisis behaviour at once. A better forecast did not do it; suppressing trades did, and the evidence showed that it is *when* trades are blocked that matters. Decision: a no-trade band that widens with benchmark volatility, now the backtester default. It adds +2.35pp of return and +0.13 of Sharpe, makes the drawdown 3.1pp shallower, and loses −2.5% through the GFC against −14.5%. It was chosen over a constant 15% band for robustness, at a cost of 0.03 Sharpe and 4.9pp of drawdown.
- **The maximum drawdown (Section 10).** The deepest drawdowns are sell-offs in the names held, not in the market, and each is dominated by one position at the cap. A stop measured from each position’s own high-water mark and sized in units of its own volatility takes the maximum drawdown from −43.5% to −27.9% and Sharpe from 1.26 to 1.41, for 1.0pp of return.
- **The live universe (Section 11).** On the twenty names the live configuration carries, at a 60% spread, the allocator has no feasible solution until SNDK’s short-history mean makes one: a spread is only reachable while it stays below the gap between the best four names’ means and the benchmark’s, and above that ceiling the book stops trading rather than degrading.
- **Universe dependence (Section 9).** Thesis: names with zero weight move the allocation through the jointly fitted DCC dynamics. Confirmed: $\bar{Q}$ is exactly invariant, α is not. Fix: fit the dynamics on the ten names with the highest mean return plus the benchmark. The influence of laggards falls from 4.75pp to 0.0002pp. In the backtest it costs nothing (33.47%, 1.33, −40.1% against 33.20%, 1.30, −42.8% for the control).

1 The DCC specification

Let $r_t \in \mathbb{R}^k$ be the vector of asset returns. Engle and Sheppard decompose the conditional covariance as

$$H_t = D_t R_t D_t, \quad D_t = \text{diag}\left(\sqrt{h_{1t}}, \dots, \sqrt{h_{kt}}\right), \quad (1)$$

where R_t is the conditional correlation matrix and each h_{it} follows a univariate GARCH process,

$$h_{it} = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip} r_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{i,t-q}. \quad (2)$$

Writing $\epsilon_t = D_t^{-1} r_t$ for the standardised residuals, the correlation dynamics are

$$Q_t = \left(1 - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n\right) \bar{Q} + \sum_{m=1}^M \alpha_m \epsilon_{t-m} \epsilon'_{t-m} + \sum_{n=1}^N \beta_n Q_{t-n}, \quad (3)$$

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}, \quad Q_t^* = \text{diag}(\sqrt{q_{11t}}, \dots, \sqrt{q_{kk t}}), \quad (4)$$

so that a typical element of R_t is $\rho_{ijt} = q_{ijt} / \sqrt{q_{iit} q_{jjt}}$.

The definition that matters for what follows is the one the paper states immediately after equation (3): $\bar{Q}$ is the **unconditional covariance of the standardised residuals** produced by the first stage,

$$\bar{Q} = \frac{1}{T} \sum_{t=1}^T \epsilon_t \epsilon'_t. \quad (5)$$

This is *correlation targeting*: it pins the long-run level of the recursion to a quantity estimated directly from the data, leaving only (α, β) to be fitted. Sufficient conditions for $H_t \succ 0$ are $\alpha_m \geq 0$, $\beta_n \geq 0$, $\sum_m \alpha_m + \sum_n \beta_n < 1$, and $\lambda_{\min}(\bar{R}) > \delta > 0$. Estimation proceeds in two stages: univariate GARCH models first, then (α, β) by quasi-maximum likelihood on the standardised residuals, maximising

$$QL_2(\psi \mid \phi, r) = -\frac{1}{2} \sum_{t=1}^T \left(\log |R_t| + \epsilon'_t R_t^{-1} \epsilon_t \right). \quad (6)$$

When $\alpha + \beta < 1$ the process mean-reverts, and multi-step forecasts of the correlation revert toward $\bar{Q}$:

$$\mathbb{E}_t[Q_{t+k}] \approx (1 - (\alpha + \beta)^{k-1}) \bar{Q} + (\alpha + \beta)^{k-1} Q_{t+1}. \quad (7)$$

2 The defects

2.1 D1: $\bar{Q}$ is computed on scrambled raw returns (0.3.0 only)

Version 0.3.0 computes the targeting matrix at three places — inside both likelihoods and inside `predict` — as

```
Q_bar = np.cov(self.rt.reshape(self.N, self.T))
```

whereas the project's GitHub master branch uses

```
Q_bar = np.cov(self.rt.T)
```

There are two independent errors in the 0.3.0 line; the second is specific to it.

Wrong quantity. `self.rt` holds de-meanned *raw returns*, not the standardised residuals ϵ_t that equation (5) requires.

Wrong operation. `np.cov` treats rows as variables, so an asset-major ($k \times T$) layout is wanted — but `reshape` is not a transpose. For a row-major ($T \times k$) buffer it reads the flat sequence $r_{11}, r_{12}, \dots, r_{1k}, r_{21}, \dots$ and cuts it into k rows of length T , so each resulting “asset” interleaves different assets at different timestamps. The covariance is computed on an array that does not correspond to any asset.

On three series with true pairwise correlation $\rho_{AB} = 0.70$, the targeting matrix itself is destroyed:

$\bar{Q}$ construction	$\bar{Q}_{AB}$	matches truth?
<code>reshape(N, T)</code> (0.3.0)	−0.0242	no
<code>.T</code> transpose (master)	0.7067	yes
standardised residuals (Engle above)	0.7078	yes

Running the two variants end to end isolates the effect. Both were built by loading the installed 0.3.0 module source and substituting only that one expression, so nothing else differs:

Forecast correlation	ρ_{AB}	ρ_{AC}	ρ_{BC}
Truth	0.70	0.30	0.20
Sample	0.7067	0.3304	0.2064
0.3.0, <code>reshape(N, T)</code>	0.1030	0.0526	0.0645
master, <code>.T</code>	0.6675	0.3039	0.1619

The reshape alone accounts for the collapse. Master’s residual error — 0.6675 against a 0.7067 sample, 0.1619 against 0.2064 — is attributable to D2 and D4, which it shares with 0.3.0.

2.2 D2: (α, β) are never estimated (both versions)

`fit()` returns $\alpha = 0.01, \beta = 0.94$ — bit-identical to the starting values passed to `scipy.optimize.minimize`. The stage-two likelihood (6) terminates without moving. With β pinned at 0.94, Q_t is an exponentially weighted average with an effective memory of $1/(1 - \beta) \approx 17$ observations, and total weight $\alpha/(1 - \beta) \approx 0.167$ on the outer-product term. The “30-day forecast” is therefore a noisy ~ 17 -day conditional estimate pulled toward the scrambled target of D1.

2.3 D3: the horizon scales the covariance by $\sqrt{h}$ (both versions)

`predict(ndays)` returns `H.t[-1] * np.sqrt(ndays)` in both 0.3.0 and master. Variance accumulates linearly in time; only *volatility* scales with $\sqrt{h}$. At the production horizon of 30 days the covariance is understated by a factor of $30/\sqrt{30} = \sqrt{30} \approx 5.48$.

This error is benign for the allocation itself — a uniform positive scalar on H leaves $\arg \min_w w' H w$ unchanged subject to the same linear constraints — but it makes every reported risk number wrong, and it interacts with the fixed 10^{-8} ridge added to H before solving.

2.4 D4: no multi-step forecast (both versions)

The package returns the last in-sample conditional covariance, scaled. It does not evaluate (7), so no mean reversion toward $\bar{Q}$ occurs over the forecast horizon.

2.5 D5: one short-history symbol truncates the whole window

Independently of `mgarch`, the caller reduced the panel with a single intersection,

```
clean_returns = asset_returns.dropna(how='any')
```

so the estimation window is the *intersection* of all listing histories. In the live universe, one 2025 spin-off with 567 prints alongside nineteen symbols with 1,829 collapsed the covariance for every asset to 567 observations — discarding 69% of the available data for 18 of 20 names. The nominal “five-year lookback” was, for the risk model, nineteen months.

3 The correction

3.1 A specification-faithful estimator

`mgarch` is replaced by an in-repository implementation (`xantos/port_opt/dcc.py`; mirrored at `algorithms/titi/dcc.py` in the backtester) that follows Section 1 directly: univariate GARCH(1,1) by QML per series; $\bar{Q}$ from standardised residuals per (5); (α, β) estimated by maximising (6)

subject to $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$; the normalisation (4); and a genuine multi-step forecast in which both the variance path and (7) mean-revert, aggregated over the horizon as

$$\widehat{\text{Cov}}\left(\sum_{k=1}^H r_{t+k}\right) = \sum_{k=1}^H D_{t+k} R_{t+k} D_{t+k}. \quad (8)$$

3.2 Per-pair estimation windows

Rather than intersecting all histories, symbols are grouped into history tiers and estimated longest tier first. Each shorter tier is then estimated jointly with every longer symbol over the window that shorter tier supports, and only the rows and columns it introduces are retained. Blocks are merged in *correlation* space — correlations are scale-free, so estimates from different windows compose — and the result is repaired to positive definiteness by flooring its eigenvalues:

$$\Sigma = \mathcal{P}(\tilde{R} \odot \sigma\sigma'), \quad \mathcal{P}(A) = V \max(\Lambda, \varepsilon \lambda_{\max}) V', \quad (9)$$

where $\tilde{R}$ carries, for each pair, the correlation estimated on the longest window both members share, and σ carries each asset’s own longest-window volatility.

For the live universe this yields three tiers — 1,827 observations for eighteen symbols, 901 for one 2024 listing, 567 for one 2025 spin-off — instead of 567 for all twenty. Where no short-history symbol is present the procedure reduces to a single estimate: verified identical to the prior single-block path to 1.1×10^{-18} absolute.

4 Evidence

4.1 Correlation recovery

Test	Truth / sample	Corrected	mgarch
Constant ρ , 3 assets	0.70 / 0.7067	0.6959	0.1030
Constant ρ , 3 assets	0.30 / 0.3304	0.3333	0.0526
Regime switch (recent $\rho = 0.9$)	0.8994	0.8956	0.2315
PANW–FTNT, 5Y daily	0.6074	0.5990	0.2157
VLO–JPM, 5Y daily	0.2970	0.2754	0.0167
VLO–PANW, 5Y daily	0.1323	0.1101	−0.0345

The regime-switch row is the most diagnostic: a two-regime series whose correlation shifts from 0.1 to 0.9 at the midpoint. A working conditional model should forecast close to the recent regime rather than the 0.474 full-sample average. The corrected estimator returns 0.8956; **mgarch** returns 0.2315, below even the unconditional value. Note also that **mgarch** reports a *negative* correlation for VLO–PANW where the sample correlation is positive.

4.2 Horizon scaling

h	Corrected (ratio to $h=1$)	Correct value	mgarch
10	10.000	10	3.162
30	30.000	30	5.477

4.3 Implied volatilities

Annualised volatility implied by the forecast, against the sample estimate:

Symbol	Sample	Corrected	mgarch
AAPL	23.2%	22.7%	6.8%
VLO	30.5%	28.3%	8.6%
PANW	35.6%	36.5%	11.4%

5 Implications

Reported risk was understated roughly five-fold. D3 alone deflated every variance by $\sqrt{30}$. Any risk figure derived from the DCC matrix — forecast portfolio volatility, risk contributions, any downstream value-at-risk — was wrong by that factor. Allocations were unaffected, because the error is a uniform scalar; reported risk was not.

The optimiser was diversifying against noise. Minimising $w'Hw$ is almost entirely a statement about off-diagonals. With correlations compressed toward zero, every asset looked like an effective diversifier, so the optimiser spread weight widely and produced portfolios whose apparent risk reduction was an artefact. With correlations restored, genuine diversification is scarcer and the solution concentrates: on the live universe at a 40% spread, holdings fall from 15 to 8.

Cash demand was an artefact. Under the defective estimator the optimiser wanted 5% in the benchmark index — which the system maps to cash, so that weight is capital left uninvested. Once correlations were corrected that demand went to zero without any change to the constraint set: the index had been attractive only because it looked like a low-variance, near-uncorrelated asset. A `max_cash_reserve_percent` bound was added as a guardrail, but it is no longer binding.

Effective lookback was a third of the intended window. D5 meant the risk model used 567 observations where 1,829 were available. Because the truncation was anchored to the newest listing’s start date rather than the configured lookback, the estimate was also insensitive to the lookback parameter: changing the five-year window moved nothing.

The 2022 adoption decision rests on the defective estimator. The August 2022 study that introduced DCC measured a peak gain of +2.08% annualised at a 30% spread. That result was produced with D1–D4 present. Section 6 re-runs the study across all three estimators.

6 Backtest replication

6.1 Design

The tables in this section were produced before the rebalance-trigger defect of Section 7 was found; the 30% point is re-based there, and the estimator conclusion is unchanged.

The August 2022 study is re-run over its original window (December 2002 to May 2022), its footnote-1 universe, monthly rebalancing, a 5% minimum trade offset, \$100,000 of capital, dividends reinvested, tax-loss harvesting disabled, a 30% position cap, BCP disabled and the S&P 500 as benchmark. Four arms isolate each defect:

Arm	Estimator	Window	Isolates
<code>gclpm</code>	SCLPM (no DCC)	–	the 2022 baseline
<code>mgarch</code>	mgarch 0.3.0	truncated	<i>as shipped</i>
<code>dcctrunc</code>	corrected	truncated	the estimator effect
<code>dcc_fixed</code>	corrected	tiered	the window effect

`mgarch` and `dcctrunc` were run from git worktrees pinned to the pre-fix commit, so the comparison is between real code paths rather than a simulated switch.

Three deviations are recorded. The study runs at 100bps rather than the repository’s current 200bps, matching the paper; the fee is identical across arms, so the deltas are unaffected. `SQ` is replaced by `XYZ` following Block’s ticker change, with history verified continuous from 2015-11-19. Regional limits are `ADR:10%` and `BND:5%`, since the paper states only that ADRs and bond ETFs were more tightly limited without giving values.

6.2 Configuration required to obtain valid results

Runs with `fractional_trading: true` – the setting the paper explicitly reports using – diverge. Fills are sized as $\text{cash_balance} \times (1 - \text{reserve}) / \text{price}$ to six decimals and consume cash to essentially zero; fees then tip the balance negative, and once portfolio value is negative `order_target_value()` computes $\text{weight} \times \text{portfolio_value} < 0$ and issues *short sales*, because `short_limit` constrains the optimiser’s weights and not execution. The position then amplifies without bound.

Evidence: at 20% spread the first negative portfolio value is 2008-11-12 and the first `SELL_SHORT` is 2008-12-19, so the negative value precedes the shorts; `transactions.json` contains genuine `SELL_SHORT` fills under `short_limit: 0.0`; final snapshots show every position at large negative quantity (ZTS at -53.6 billion shares); and `Net-Fees P/L (%)` reaches -7504% , which is impossible for a long-only book. The runs complete successfully and write plausible-looking statistics throughout.

Integer share sizing removes this: rounding each fill down leaves a residual buffer on every trade. Validated on the three configurations that diverged worst, over the full window: spreads 5%, 20% and 27.5% all return zero negative-value days and zero short fills, with annualised returns of 16.2%, 24.7% and 27.8%. A 20bps cash reserve alone was *not* sufficient – it moved the first failure at 15% spread from 2006-08-22 to 2017-06-27 without preventing it. All results below therefore use integer share sizing, which is a deviation from the paper and is discussed in the caveats below.

6.3 The 2022 result replicates under the shipped estimator

The paper reports a peak advantage of $+2.08\text{pp}$ annualised at a 30% spread and a -1.81pp deficit at the conservative end. The `mgarch` arm – the code that actually ran in 2022 – reproduces both: $+3.48\text{pp}$ at exactly 30% and -2.03pp at 5%, with the same sign change near 10%. The environment therefore reproduces the study.

6.4 The advantage does not survive correcting the estimator

At the paper’s own peak spread of 30%, the advantage over the baseline falls from $+3.48\text{pp}$ to $+1.06\text{pp}$ when only the estimator is corrected, and to $+0.13\text{pp}$ when the estimation window is also corrected. Attributing the 3.35pp that disappears: the estimator defects account for 2.42pp (72%) and the window truncation for 0.92pp (28%).

Across spreads of 15% and above the corrected arm’s deltas are $+0.06$, -0.38 , $+0.04$, -0.41 , $+0.13$ and $+2.10\text{pp}$ – six values straddling zero. There is no systematic return advantage to DCC over the lower-partial-moment baseline once the estimator is specification-faithful.

6.5 What the shipped estimator was actually doing

Two measurements explain the illusory edge. First, the defective covariance implied a mean pairwise correlation far below sample, so every asset appeared to be a diversifier: the `mgarch` arm holds 20.9 names with an effective breadth of 12.0, against 9.2 names and effective breadth

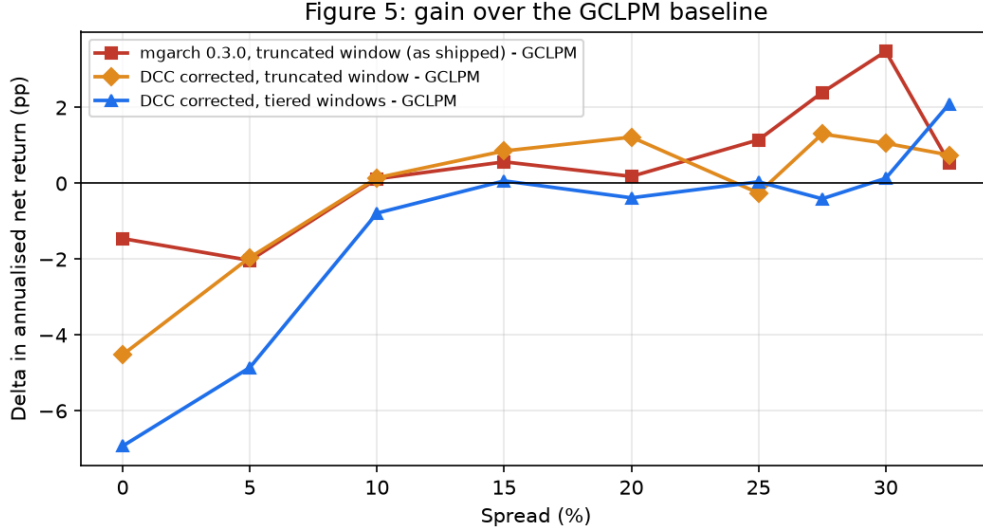


Figure 1: Gain in annualised net return over the GCLPM baseline, by spread. The shipped estimator reproduces the 2022 peak at 30%; the corrected estimator does not.

4.8 for the corrected arm. Second, that over-diversified book is well suited to trending markets and poorly suited to drawdowns, which is exactly the pattern in the year-by-year returns.

At a 15% spread the **mgarch** arm beats the baseline by 14.8pp in 2013, 13.3pp in 2020, 8.7pp in 2007 and 7.9pp in 2017, and loses by 7.8pp in 2018 and 9.7pp in 2022. Its outperformance is concentrated in trending years and reverses in stress years. The risk it took is visible directly: at 30% spread its maximum drawdown is -52.3% against -46.0% for the baseline and -46.3% for the corrected arm, and through the GFC crash it lost -27.3% against -14.5% and -18.9% .

6.6 Crisis behaviour

Through the 2022 selloff at 15% spread the corrected arm lost -9.5% , the baseline -10.4% , the estimator-only arm -14.4% and the shipped estimator -20.1% : the corrected estimator lost less than half what the defective one did. The COVID crash separates the arms by less than 0.5pp at 30% spread (-28.0% to -28.4%), so no estimator offered protection in that episode.

6.7 What survives

The baseline's advantage at conservative spreads replicates and is large: at 0% spread it returns 21.18% annualised with a Sharpe of 1.69 and a -25.4% maximum drawdown, against 14.25%, 1.22 and -34.6% for the corrected DCC arm. The paper's own finding that the lower-partial-moment approach is preferable for conservative mandates is confirmed, and by a wider margin than it reported.

6.8 Caveats

These are single-path backtests. With twenty annual observations on a concentrated 34-name universe, differences of roughly one percentage point are not distinguishable from noise; the results relied on above are the large and consistent ones – the conservative baseline advantage and the collapse of the 30% peak. Integer share sizing departs from the paper's stated configuration, which could not be run to completion. Tiingo history for this universe begins in 2010 for the supplementary correlation analysis, so those figures exclude the GFC. No statistical test of the differences has been performed.

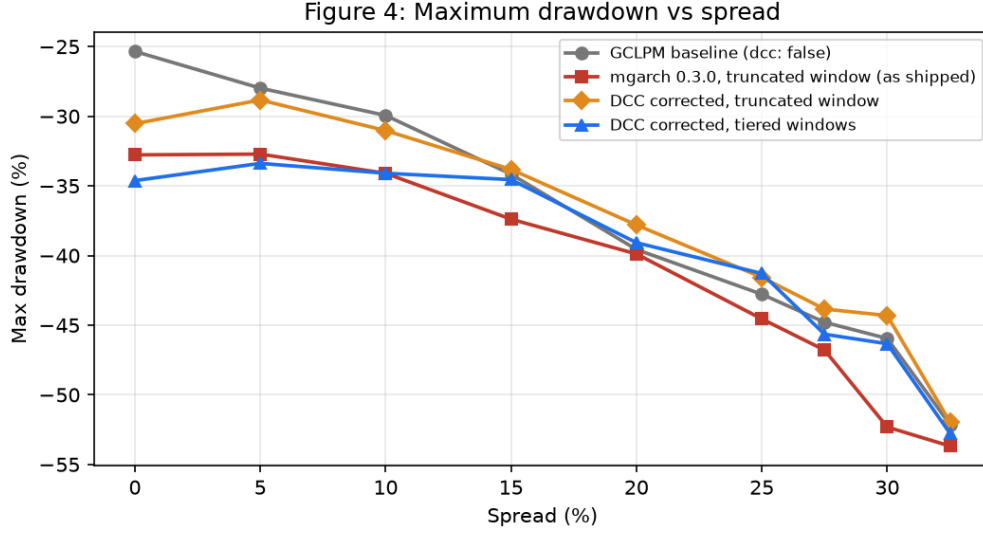


Figure 2: Maximum drawdown by spread. The shipped estimator is deepest at every spread from 0% to 30%.

6.9 Full results

6.10 Annualised net return

Spread	GCLPM	mgarch 0.3.0	DCC fixed, trunc.	DCC fixed, tiered
0.0%	21.18%	19.72%	16.66%	14.25%
5.0%	21.06%	19.03%	19.10%	16.20%
10.0%	21.54%	21.66%	21.68%	20.75%
15.0%	23.69%	24.26%	24.55%	23.76%
20.0%	25.13%	25.31%	26.34%	24.74%
25.0%	26.70%	27.84%	26.44%	26.74%
27.5%	28.23%	30.63%	29.54%	27.83%
30.0%	30.85%	34.33%	31.91%	30.98%
32.5%	30.18%	30.72%	30.93%	32.28%

6.11 Annualised net volatility

Spread	GCLPM	mgarch 0.3.0	DCC fixed, trunc.	DCC fixed, tiered
0.0%	12.51%	13.57%	11.96%	11.72%
5.0%	13.18%	13.73%	12.71%	12.31%
10.0%	13.61%	14.89%	14.02%	13.64%
15.0%	15.78%	16.53%	15.65%	15.51%
20.0%	18.84%	19.22%	18.01%	18.45%
25.0%	22.67%	22.99%	22.19%	22.16%
27.5%	24.21%	24.48%	23.66%	24.21%
30.0%	26.57%	26.52%	25.18%	25.71%
32.5%	27.68%	27.19%	27.48%	27.65%

6.12 Sharpe ratio

Spread	GCLPM	mgarch 0.3.0	DCC fixed, trunc.	DCC fixed, tiered
0.0%	1.69	1.45	1.39	1.22
5.0%	1.60	1.39	1.50	1.32
10.0%	1.58	1.45	1.55	1.52
15.0%	1.50	1.47	1.57	1.53
20.0%	1.33	1.32	1.46	1.34
25.0%	1.18	1.21	1.19	1.21
27.5%	1.17	1.25	1.25	1.15
30.0%	1.16	1.29	1.27	1.21
32.5%	1.09	1.13	1.13	1.17

6.13 Maximum drawdown

Spread	GCLPM	mgarch 0.3.0	DCC fixed, trunc.	DCC fixed, tiered
0.0%	-25.3%	-32.8%	-30.5%	-34.6%
5.0%	-28.0%	-32.7%	-28.8%	-33.4%
10.0%	-29.9%	-34.1%	-31.0%	-34.1%
15.0%	-34.2%	-37.4%	-33.8%	-34.6%
20.0%	-39.6%	-39.9%	-37.8%	-39.1%
25.0%	-42.8%	-44.6%	-41.6%	-41.3%
27.5%	-44.8%	-46.8%	-43.8%	-45.7%
30.0%	-46.0%	-52.3%	-44.3%	-46.3%
32.5%	-52.1%	-53.7%	-51.9%	-52.8%

7 A defect in the backtester's rebalance trigger

7.1 What it was

The backtester placed orders only when the optimiser's target had changed, and tested that with

```
fast_path = (self.last_allocations - self.allocations).sum() != 0
```

Both vectors are weights that sum to one, so their difference sums to zero up to floating-point noise *whatever* the change. A swap of two names at the same weight — one 30% holding replaced by another, the most common move this allocator makes — cancels exactly, and the rebalance was skipped in silence. Whether any trading happened at all came down to rounding.

7.2 Evidence

Reproducing the test on the logged targets predicts 87% of the trade/no-trade outcomes across 152 rebalances of the corrected-DCC arm: where the sum was exactly zero the book did not trade in 43 of 55 cases, and where it was non-zero the book traded in 89 of 97. The consequences in that run:

- 39 target changes of 10 percentage points or more — up to 40pp — never reached the market. No order was rejected and nothing was logged; the orders were never placed.
- The book sat more than 20% away from its target on 32% of days, with a 90th percentile distance of 43% one-way. After the fix that falls to 15%.

- Positions drifted far above the 30% cap, because a trim that the optimiser wanted was never sent: NFLX reached 54% of the book in 2013, BKNG and MKTX 47%.

The live rebalancer (`xantos/rebalance.py` in `xantos-ra`) has no such gate: it walks the target allocation and applies the no-trade band per order. Live trading therefore acted on every target change while the backtests acted on about two thirds of them, so the backtests were not simulating the live system.

7.3 The fix

```
def allocations_changed(last, new, tol=1e-9):
    diff = last.sub(new, fill_value=0.0).abs()
    return bool(diff.max() > tol) if len(diff) else False
```

Any single weight moving more than the tolerance triggers the rebalance; the no-trade band then decides, per order, what is worth trading. Five unit tests cover it, including the equal-weight swap that used to cancel.

7.4 Impact

Every arm was re-run at the 30% spread. Trade counts rise by 17 to 25%, which is the defect’s direct signature. The first four numeric columns are the runs as they were made, with the defect present; the last four are the same arms re-based on the fixed engine.

Arm, 30% spread	Ret.	Sharpe	Max DD	Trades	Ret.	Sharpe	Max DD	Trades
GCLPM baseline	30.85%	1.16	-46.0%	576	30.59%	1.16	-45.1%	710
mgarch 0.3.0 as shipped	34.33%	1.29	-52.3%	792	32.70%	1.23	-47.4%	959
DCC corrected, truncated	31.91%	1.27	-44.3%	647	30.81%	1.20	-48.2%	793
DCC corrected, tiered	30.98%	1.21	-46.3%	625	29.37%	1.12	-46.4%	776
+ state band + top-10	33.47%	1.33	-40.1%	595	32.35%	1.26	-43.5%	733

The advantage over the baseline, which is what Section 6 concludes from:

Annualised return vs GCLPM baseline	as run	re-based
mgarch 0.3.0 as shipped	+3.48pp	+2.11pp
DCC corrected, truncated window	+1.06pp	+0.21pp
DCC corrected, tiered windows	+0.13pp	-1.22pp
+ state band + top-10 reference	+2.62pp	+1.75pp

The estimator conclusion is unchanged and slightly sharper. The shipped estimator still beats the baseline by a wide margin at the paper’s peak spread, and correcting it still gives that advantage back — now more than all of it, at -1.22pp . The 2022 adoption result remains an artefact of the defect.

7.5 Scope and what it does not tell us

Only the 30% spread has been re-based; the nine-spread tables in Section 6 and the arm comparisons in Section 8 were produced before the fix. Re-running the full sweep is 36 backtests, so the 30% point — the paper’s peak and the only spread at which its headline is claimed — was re-based first.

A second lesson bears on how all of these comparisons should be read. A block bootstrap of the monthly differences between two arms over this window gives a 95% interval for the

annualised difference of roughly $\pm 3\text{pp}$. Differences smaller than that are not resolvable on a single path, whatever the point estimate says. The band variants of Section 8 differ from one another by less than that, and the apparent knife-edge behaviour of the constant band is consistent with tie-breaking noise: at a 30% spread the optimiser’s target is usually three names at the position cap plus one small holding, and which names occupy the capped slots is near-degenerate — a perturbation of 0.5% in the covariance is enough to swap them.

8 Crisis behaviour: searching for a DCC change that works

8.1 Goal and the gap

Every arm in this section was measured before the rebalance-trigger defect of Section 7 was found. The re-based 30% figures are in that section: the state band keeps its advantage over the baseline (+1.75pp of return, +0.10 Sharpe, 1.6pp shallower drawdown) but it is smaller than measured here, and the differences between the individual band variants below are inside the $\pm 3\text{pp}$ that a single path can resolve.

The goal set for this stage was a change to the DCC allocator that beats the GCLPM baseline on annualised return, on Sharpe ratio *and* on crisis behaviour at the same time. Everything in this section is at the 30% spread, the configuration the 2022 study adopted. Crisis behaviour is measured by maximum drawdown and by the return through three windows: the GFC (2007-10-09 to 2009-03-09), the COVID crash (2020-02-19 to 2020-03-23) and the 2022 selloff (2022-01-03 to 2022-05-31). All figures are recomputed from each run’s portfolio value series by `docs/analyze_crisis_study.py`.

The corrected estimator clears the first two goals narrowly and fails the third: +0.13pp of return and +0.04 of Sharpe, but a maximum drawdown 0.38pp deeper than the baseline, a GFC loss of -18.9% against -14.5% , and -32.5% against -31.3% in 2022. The gap to close is crisis behaviour.

Arm (30% spread)	Return	Sharpe	Max DD	Trades	GFC	COVID	2022
GCLPM baseline	30.85%	1.16	-46.0%	576	-14.5%	-28.4%	-31.3%
mgarch 0.3.0 as shipped	34.33%	1.29	-52.3%	792	-27.3%	-28.0%	-35.4%
Corrected DCC	30.98%	1.21	-46.3%	625	-18.9%	-28.4%	-32.5%
DCC, correlation shrink 0.25	32.65%	1.23	-46.9%	593	-19.8%	-25.6%	-31.3%
GARCH diagonal + SCLPM	27.79%	1.04	-48.9%	615	-25.0%	-28.4%	-35.3%
Turnover penalty, $g=1$	31.42%	1.19	-45.4%	550	-18.1%	-29.2%	-30.8%
Turnover penalty, $g=5$	29.98%	1.13	-52.0%	487	-27.3%	-29.3%	-33.9%
Constant band 0.10	29.79%	1.17	-47.4%	499	-17.0%	-29.4%	-31.7%
Constant band 0.15	32.47%	1.33	-37.9%	442	-9.4%	-26.5%	-27.4%
Constant band 0.25	28.76%	1.16	-41.8%	364	-4.2%	-31.0%	-27.4%
State band, drawdown signal	32.39%	1.25	-43.7%	532	-1.3%	-30.6%	-32.6%
State band 0.05 to 0.40	33.17%	1.30	-42.8%	595	-0.8%	-29.7%	-32.2%
State band 0.05 to 0.25	33.20%	1.30	-42.8%	602	-2.5%	-29.7%	-32.2%

8.2 Thesis 1: the covariance forecast is the bottleneck

Thesis. If the crisis gap comes from the forecast itself – correlations reacting too slowly, or too noisily, as a crash develops – a better forecast should close it.

Test. Asymmetric DCC, downside-only correlation, correlation shrinkage, a 90-day horizon and Ledoit–Wolf shrinkage were screened first; the fast screening harness moved each by roughly 0.1pp, but it covered only 2015 to 2022 and so contained no GFC, and it was judged not

Figure 6: every 30%-spread arm against the GCLPM baseline

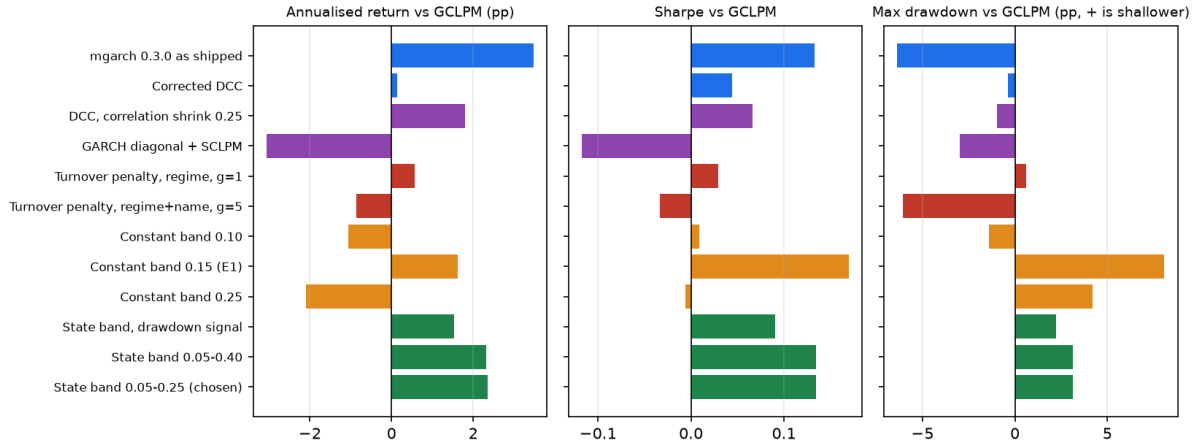


Figure 3: Every 30%-spread arm against the GCLPM baseline on the three goals. Only the constant 0.15 band, the three state-band variants and (marginally) the $g = 1$ turnover penalty are to the right of zero on all three panels.

predictive. Two variants were therefore run over the full window. Shrinking the conditional correlation 25% toward $\bar{Q}$ added 1.67pp of return over the corrected DCC but deepened the drawdown (-46.9%) and the GFC loss (-19.8%). The extension the 2022 paper itself proposed – GARCH variances on the diagonal of the SCLPM matrix – lost 3.06pp to the baseline with a -48.9% drawdown.

Conclusion: rejected. Changing the forecast moves the return; it does not change what the book does in a crash.

8.3 Thesis 2: the book trades too much

Thesis. The DCC arms trade more than the baseline – 625 trades against 576, and 792 for the shipped estimator, which also had the worst drawdown. If turnover itself destroys value, suppressing it should help.

Test. Raising the minimum trade offset from the production 5% to 15% beats the baseline on all three goals at once: +1.62pp of return, +0.17 Sharpe and a drawdown 8.1pp shallower, with a GFC loss of -9.4% on 442 trades.

But the result is a knife-edge. At a 10% band return is 1.06pp below the baseline, and at 25% it is 2.09pp below. The optimiser-side form of the same thesis, a quadratic turnover penalty, is no better: with $g = 1$ and regime scaling it gains only +0.57pp, +0.03 and +0.6pp; with $g = 5$ and per-name scaling it loses 0.87pp and falls -27.3% through the GFC, as badly as the shipped estimator. If turnover were the problem, less of it would be monotonically better. It is not: the two lowest-turnover arms, 364 trades and 487 trades, both lose.

Conclusion: partially right. Suppressing trades can help a great deal, but *how many* trades are suppressed does not predict the outcome.

8.4 Refined thesis: it is when the book trades

In a drawdown the conditional covariance spikes and correlations converge, so the optimiser wants to reposition the whole book – into the worst prices, on its least reliable estimates. A band blocks exactly those trades. A quadratic penalty has the wrong shape for this: it taxes every trade by its square, so it suppresses useful calm-period rebalancing and barely dents the wholesale repositioning a crash provokes. A hard threshold does the opposite.

The data support this. The table splits the monthly return difference between the constant 0.15 band and the corrected DCC by the state of the market at the start of each month. The thesis was first tested against the portfolio’s own volatility and drawdown (left); the implementation keys off the benchmark’s, an exogenous signal the book cannot feed back into, so the right-hand columns are the relevant evidence. Both show the same shape: the band earns its keep in stress and is a slight drag when markets are calm.

State at month start	own n	own mean	own win	bench n	bench mean	bench win
volatility above 30%	41	+0.44pp	61%	11	+1.17pp	82%
volatility 20% to 30%	76	+0.20pp	59%	30	+0.48pp	70%
volatility below 20%	114	-0.14pp	46%	191	-0.06pp	49%
drawdown deeper than 20%	20	+0.60pp	60%	26	+0.39pp	62%
drawdown 10% to 20%	31	+0.15pp	58%	21	-0.18pp	48%
drawdown under 10%	181	+0.00pp	51%	185	+0.06pp	52%

In the eight worst months for the corrected DCC the band won all eight: 2008-10 (+4.58pp), 2022-04 (+0.24pp), 2004-07 (+5.02pp), 2022-01 (+6.17pp), 2008-09 (+3.79pp), 2018-10 (+0.75pp), 2004-08 (+1.77pp) and 2020-03 (+2.22pp). Against the benchmark it keeps 99% of the corrected DCC’s upside capture (155.8% against 157.3%) while cutting downside capture from 72.0% to 63.0%, and its downside beta is 0.92 against 1.09.

The implication is that the band should be a function of market state rather than a constant: wide in stress, where it earns, and at the production value when calm, where it costs.

8.5 The state-dependent band

With σ_t the benchmark’s trailing 21-session realised volatility, annualised,

$$s_t = \min\left(1, \max\left(0, \frac{\sigma_t - \sigma_{\text{calm}}}{\sigma_{\text{stress}} - \sigma_{\text{calm}}}\right)\right) \quad (10)$$

$$b_t = b_{\text{calm}} + (b_{\text{stress}} - b_{\text{calm}}) s_t \quad (11)$$

where b_t is the minimum weight change for a trade to be placed. The parameters were read off the bucket boundaries above rather than optimised: $\sigma_{\text{calm}} = 20\%$ (below it the band is a drag), $\sigma_{\text{stress}} = 30\%$ (above it the band clearly earns), b_{calm} equal to the configured minimum trade offset (5% in production) and $b_{\text{stress}} = 25\%$. Volatility was chosen over drawdown because it leads rather than lags: the aim is to be restrictive as a crash develops, not after the trough. Drawdown was run as the robustness check.

Over 2003 to 2022 the band sits at its calm value in 79.7% of sessions, is partly raised in 14.0% and fully raised in 6.0%. Its mean was 0.139 through the GFC, 0.192 through the COVID crash and 0.116 through the 2022 selloff (Figure 9).

8.6 Result

All three state-band variants beat the baseline on all three goals. It is the only family in which every parameterisation tried does so:

vs GCLPM baseline	Return	Sharpe	Max DD	GFC
State band 0.05 to 0.25, volatility	+2.35pp	+0.13	+3.1pp	-2.5%
State band 0.05 to 0.40, volatility	+2.32pp	+0.13	+3.1pp	-0.8%
State band 0.05 to 0.25, drawdown	+1.54pp	+0.09	+2.2pp	-1.3%
Constant band 0.15	+1.62pp	+0.17	+8.1pp	-9.4%
Constant band 0.10	-1.06pp	+0.01	-1.4pp	-17.0%
Constant band 0.25	-2.09pp	-0.01	+4.2pp	-4.2%

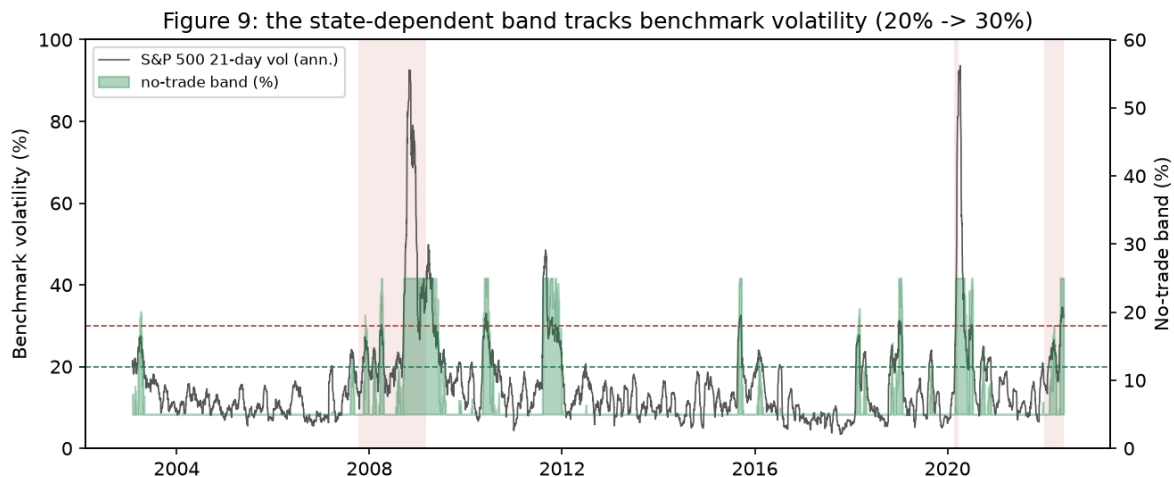


Figure 4: Benchmark volatility and the band it implies, reconstructed daily from benchmark closes. In the backtest the band is re-evaluated from the prices the allocator last refreshed, so it moves once per 30-session rebalance.

Against the corrected DCC it modifies, the chosen band adds 2.22pp of return and 0.09 of Sharpe, makes the maximum drawdown 3.5pp shallower, cuts the GFC loss from -18.9% to -2.5% , lowers downside capture from 72.0% to 63.9% and downside beta from 1.09 to 0.91.

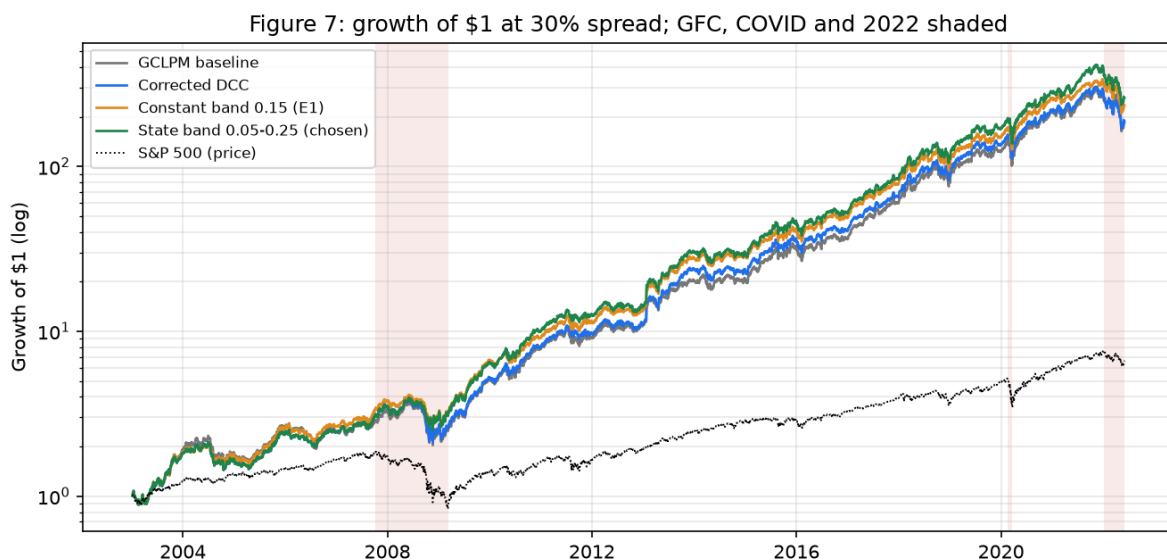


Figure 5: Growth of \$1 at a 30% spread, log scale. More than half of the state band's cumulative edge over the corrected DCC (57% in log terms) was earned in the GFC window; the COVID crash gave a little of it back.

8.7 Testing the thesis directly

If the thesis is right, the state band should beat the constant band in calm months (it no longer pays the drag) and match it in stress (both are wide). Splitting the monthly difference between the two by benchmark state:

State band minus constant 0.15	n	mean	win	sum
benchmark volatility below 20%	191	+0.11pp	52%	+21.5pp
benchmark volatility 20% to 30%	30	-0.37pp	23%	-11.0pp
benchmark volatility above 30%	11	+0.41pp	64%	+4.5pp

The first and last rows confirm the thesis: the calm-market drag is gone and nothing is given up in acute stress. The middle row is a failure the thesis did not predict. In the ramp zone, where benchmark volatility is between 20% and 30% and the band is part-way raised, the state band loses to the constant band in 23 of 30 months. The 2022 selloff was a slow decline that lived in this zone – the band averaged 0.116 – and it is where the state band gives back the constant band’s advantage: -32.2% against -27.4% . Monthly differences between two paths are path-dependent, so the row-level attribution is indicative rather than exact.

Figure 8: monthly return difference by benchmark state at the start of the month

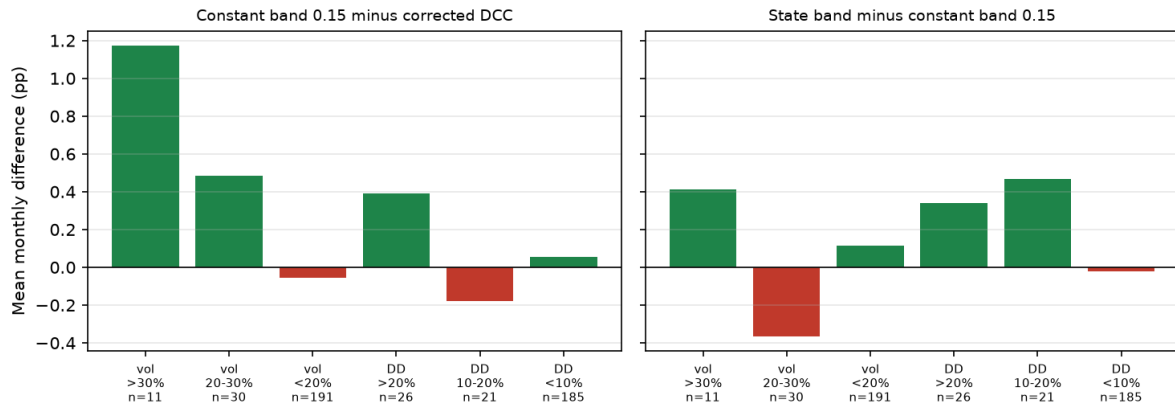


Figure 6: Left: the constant band earns in stress and costs when calm. Right: the state band removes the calm-market cost but underperforms the constant band in the 20% to 30% ramp zone.

8.8 Why the state band was chosen, and what it costs

The state band does not dominate the constant 0.15 band:

	Constant 0.15	State band	Difference
Annualised return	32.47%	33.20%	+0.73pp
Sharpe ratio	1.33	1.30	-0.03
Maximum drawdown	-37.9%	-42.8%	-4.9pp
GFC	-9.4%	-2.5%	+6.9pp
COVID	-26.5%	-29.7%	-3.2pp
2022	-27.4%	-32.2%	-4.8pp
Trades	442	602	+160

It was chosen for four reasons.

- **Robustness.** The constant band’s result exists at one value: 0.10 and 0.25 both lose to the baseline. Every state-band variant tried beats the baseline on all three goals. A setting that has to be exactly right in-sample is the less likely to be right out of sample.

- **It leaves production behaviour alone most of the time.** In 80% of sessions the book trades at the existing 5% band; the change acts only when the benchmark is volatile. A constant 15% band changes behaviour everywhere.
- **Return.** It has the highest return of any arm that beats the baseline on all three goals.
- **Mechanism.** It targets what the evidence identifies: trading when the covariance forecast is least trustworthy.

The cost is real: 0.03 of Sharpe and 4.9pp of maximum drawdown against the constant band, concentrated in slow, moderate-volatility declines that a volatility trigger does not flag. Between the two volatility variants, 0.25 was preferred to 0.40 because they are indistinguishable (33.20% against 33.17%, identical drawdown) – the band is at its maximum in only 6% of sessions – and the smaller setting is the smaller extrapolation. The drawdown signal is lower on return, Sharpe and drawdown than the volatility signal.

8.9 Decision

The state-dependent band is now the default in the backtester’s Titi algorithm: `band_stress_enabled: true, band_calm inherits min_offset_for_trade_percent, band_stress: 0.25, band_vol_calm: 0.20, band_vol_stress: 0.30, band_signal: vol`. Every earlier study’s stored configuration pins `band_stress_enabled: false` so it still reproduces. The live allocator in xantos-ra is unchanged; adopting the band for live trading is a separate decision.

The ramp zone is the open item. Starting the ramp lower, or holding a constant floor above the calm value, are the obvious candidates, but either would be tuned on the same path that scores it and should be validated out of sample – June 2022 onward – before adoption.

8.10 Caveats

The thresholds were read from the 2003 to 2022 path that also scores the result, so the win is partly in-sample; they were taken from bucket boundaries rather than optimised, and the honest reading is “the mechanism is real” rather than “these are the right numbers”. Only 11 months began with benchmark volatility above 30%. The band moves once per rebalance, so in a crash as fast as COVID (23 sessions peak to trough) it can be late. These remain single-path backtests.

9 Universe dependence

9.1 The observation

The DCC allocation depended on names the optimiser did not hold. Adding five names to a ten-name universe, each receiving zero weight, moved the weights on the original ten by 4.75pp in total (1.28pp on a single name), at a 20% spread over the five years to September 2026.

9.2 Diagnosis from first principles

Every ingredient of the covariance for a pair (i, j) was checked for dependence on the other columns.

- **Univariate GARCH.** Each variance is fitted on its own series: invariant.
- $\bar{Q}_{ij}$. The sample covariance of the two standardized residual series: invariant. Measured on the original ten-name block with and without the added names, the maximum difference is exactly 0.

- (α, β) . Two scalars fitted by one likelihood over *every* column: not invariant. Adding the five names cut α from 0.0125 to 0.0073, a 42% fall, while persistence $\alpha + \beta$ barely moved (0.982 to 0.983).

α is how strongly the latest co-movement updates every pair’s correlation. Names that barely co-move with the rest – which is also why the optimiser gives them nothing – pull the jointly fitted α down, so the conditional correlation of every pair in the book becomes less responsive to recent news. Zero weight is not zero influence.

9.3 Options considered

- **Fix (α, β) as constants.** Removes the dependence entirely but abandons estimation of the dynamics.
- **Composite likelihood over pairs.** Averages the pairwise likelihoods; the average still runs over the whole universe, so it only dilutes the problem.
- **Pair-specific (α, β) .** Invariant, but the resulting matrix is no longer guaranteed positive definite.
- **Fit (α, β) on a reference set,** then run the recursion and forecast for every name with those values. Invariant to any name outside the set, keeps estimation, and keeps positive definiteness.

9.4 The fix: fit the dynamics on the top ten

The reference set is the ten names with the highest mean return over the optimiser’s own five-year window, plus the benchmark. Defining it by performance rather than by an arbitrary list was a deliberate choice: the portfolio is built around its winners, so it is acceptable for winners to shape the dynamics and not for laggards to. Ranking on the five-year window keeps the set stable – in the 2010 to 2022 sample it retained 92% of its names month to month (0.78 changes a month), against 1.7 changes on a twelve-month window. The selection effect of fitting on winners is small: on the last window of that sample, using the winners’ (α, β) instead of the full universe’s changes the forecast correlation of the average pair by 0.007 (0.060 at most).

To stop a name hovering at tenth place from flipping the set every rebalance, selection uses an index-style buffer: any name ranked eighth or better is in; remaining places go to last period’s members still ranked twelfth or better; any gap is filled by rank. The live allocator is instantiated per run and so has no memory – there it is a plain top ten.

Configuration: `dcc_reference_top_n: 10` in both repositories (off when absent). The estimator change is `dcc_forecast_covariance(..., reference=...)` in `dcc.py`, which is now identical in both repositories.

9.5 Evidence

The fix does what it was built to do for laggards: 4.75pp of movement becomes 0.0002pp. It also makes the trade-off explicit. A name that ranks in the top ten by mean return but still receives zero weight (here CAMT and URI, ranked ninth and tenth) now displaces a member of the reference set, and moves the allocation *more* than before – 2.89pp against 1.28pp. That is the “winners may skew, losers may not” rule working as intended, but it means the fix makes the allocation invariant to laggards, not to every addition.

Figure 11: allocation change caused by adding names that receive zero weight

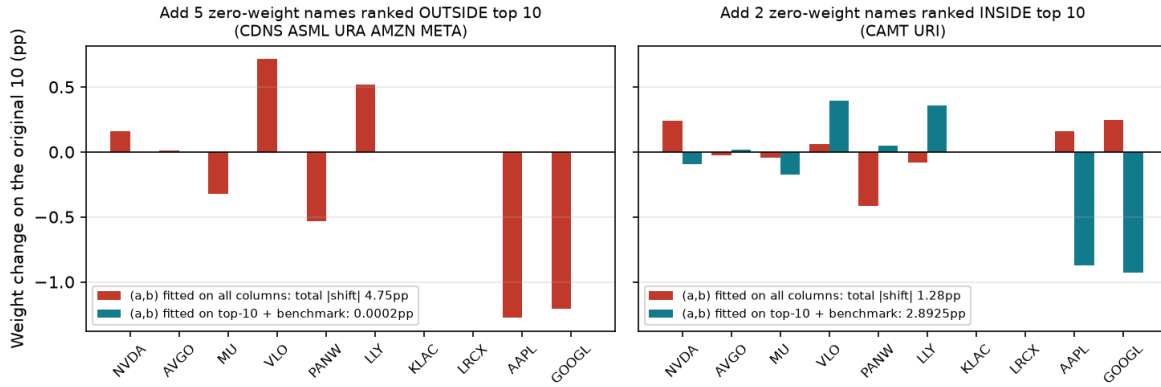


Figure 7: Weight change on the original ten names when zero-weight names are added. Left: five names ranked outside the top ten – the change falls from 4.75pp to 0.0002pp, which is solver tolerance. Right: two names ranked ninth and tenth – they enter the reference set by design and the change rises from 1.28pp to 2.89pp.

9.6 Backtest

Four runs at a 30% spread with the state band on (now the default): a control, the control with `dcc_reference_top_n: 10`, and each of those with five names added to the universe (GE, C, AIG, F and XRX). These four also predate the rebalance-trigger fix of Section 7; the re-based figure for the control’s configuration is 32.35% return, 1.26 Sharpe and a -43.5% drawdown. The invariance result below is unaffected, because it compares optimiser targets, which the trigger never touched.

Arm (30% spread)	Return	Sharpe	Max DD	Trades	GFC	COVID	2022
Control (state band default)	33.20%	1.30	-42.8%	602	-2.5%	-29.7%	-32.2%
Top-10 reference	33.47%	1.33	-40.1%	595	-2.5%	-29.7%	-28.3%
Control + 5 names	32.54%	1.26	-44.3%	621	+2.8%	-29.7%	-33.2%
Top-10 reference + 5 names	32.20%	1.24	-43.5%	618	-3.8%	-29.7%	-32.2%

The default is ported faithfully. The control reproduces the Section 7 state-band run exactly: a daily weight gap of 0.0000pp and identical return.

The reference set costs nothing. Against the control it adds 0.27pp of return and 0.03 of Sharpe and makes the maximum drawdown 2.7pp shallower, almost all of it in 2022 (-28.3% against -32.2%); GFC and COVID are unchanged. On one path this reads as “no cost”, not as an edge. The set is stable: 0.56 names change per rebalance, and it changes at 48% of rebalances.

The five added names were not laggards throughout. They were chosen for poor returns over 2002 to 2022 as a whole, but in rolling five-year windows several were not: C held up to 18% of the book in 2003 and 2004, XRX 13% in 2007 and F up to 21% in 2013 and 2014, and one of them sat in the reference set on 27% of days (AIG and C after their 2009 rebound). On those days the allocation should change, so the full-sample comparison does not test invariance.

Where the names were irrelevant, the fix works. Restricting to the 3,663 days (72%) on which no added name was held by either arm or in the reference set, the books with and without the five names are identical on 92.4% of days with the reference set (mean gap 0.58pp), against 81.4% (1.22pp) without it. These series are the optimiser’s target allocations, not holdings, so the comparison is of what the allocator asked for rather than what the book held. The

remainder is mostly carry-over: with the reference set the share of identical days rises from 82% in the 60 sessions after an episode to 93–94% thereafter, because the buffer rule remembers the previous set and a laggard’s spell inside it keeps the two sets apart for several rebalances afterwards. Without the reference set the share stays between 76% and 86% however long ago the last episode was – the signature of a persistent universe effect. The rebalance-to-date mapping recovered 152 of 163 rebalances, and some of the residual is likely from those and from the buffer rule’s memory.

One residual is not explained by the DCC dynamics. On 2005-02-28 both arms without the added names swapped a capped 30% position from TMO to PEG while both arms with them kept TMO, and all four reversed at the next rebalance. In the reference arm no added name was held or in the reference set, so (α, β) and $\bar{Q}$ on the original names were unchanged. The likeliest cause is a near-tie between the two names at the cap, which any small numerical difference decides; this has not been verified.

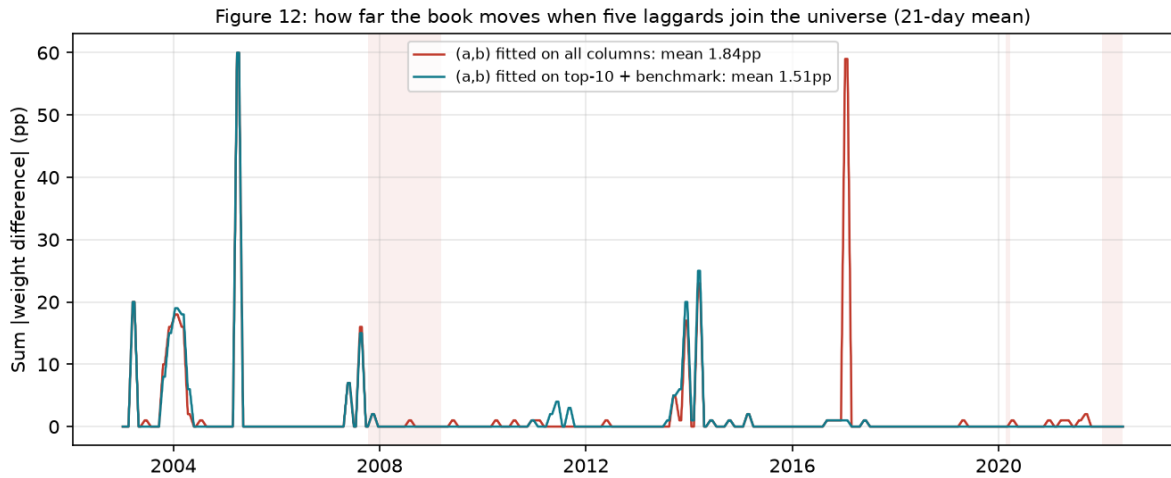


Figure 8: Daily weight gap between each arm with and without the five added names (21-day mean). The spikes in 2003–04, 2007 and 2013–14 are periods when an added name was held or in the reference set. The December 2016 to January 2017 spike is in the control arm only, with no added name held: the universe effect the fix removes. The 2005 spike in both arms is a single swap of one capped 30% position (TMO for PEG), reversed at the next rebalance.

Recommendation. Adopt `dcc_reference_top_n: 10` alongside the state band. It removes the dependence it was built to remove, at no measured cost. It is not yet the code default in either repository; making it one is a one-line change once agreed.

10 Reducing the maximum drawdown

10.1 Where the drawdown comes from

On the re-based runs the worst drawdown is not a market crash. The combined arm’s deepest episode runs from 2021-11-16 to 2022-05-11 and costs 43.5% while the S&P 500 falls 15.8%; 2006 costs 34.9% against a flat index and 2004 costs 33.5% against -2.9% . In the GFC, by contrast, the book loses less than the index. The drawdowns that matter are sell-offs in the names held, not in the market.

Reconstructing the actual holdings from the fills, each episode is dominated by one position at or near the position cap: NFLX cost 32pp of the 2004 episode, CRM 16pp of 2006, TEAM 20pp of 2022 and MA 22pp of the GFC. Two further measurements separate the causes. Comparing the correlation the model forecast at the start of each episode with what was realised

inside it:

Episode	Forecast correlation	Realised	Ratio
2022	0.32	0.55	1.71
GFC	0.26	0.45	1.73
2004	0.16	0.14	0.89
2006	0.18	0.20	1.12

In 2022 and the GFC the risk model under-forecast co-movement by about 70%: the optimiser believed it was diversified and was not. In 2004 and 2006 the correlations were right and a single name simply collapsed. A constraint keyed to the DCC correlations – a cap on the combined weight of a correlated cluster – cannot fix the first case, because it reads the same correlations that fail; measured in the harness it never binds, and its results are identical to the base to two decimal places.

10.2 What was tried

A harness replays the optimiser at all 154 rebalances and holds each target to the next, which costs seconds per variant once the covariance is cached, against four hours for a backtest. Position-level and portfolio-level constraints were screened there. Capping each name by its own volatility helped the idiosyncratic episodes and hurt the correlated ones; cutting the return target when the top holdings are falling improved the whole drawdown distribution (mean of the five deepest: -31.4% against -36.7%) at a cost of 1.5pp of return; an ex-ante exposure target did the same more expensively. In the backtester, a tighter 20% position cap made the drawdown *worse* (-45.7% against -43.5%): five names of the same factor are still one bet.

10.3 The change: a volatility-scaled trailing stop

The rule that worked is a stop on each position, measured from its own high-water mark and sized in units of its own volatility:

$$\text{stop}_i = \text{clip}(k \cdot \sigma_i^{(21)} \sqrt{21}, 8\%, 35\%), \quad k = 2.5 \quad (12)$$

$$\text{exit when } P_{i,t} \leq (1 - \text{stop}_i) \cdot \max_{s \leq t} P_{i,s} \quad (13)$$

with the position sold to cash, and no re-entry until the price is back above its 21-day average (minimum wait 21 sessions). Three design points, each forced by the data:

- **From the peak, not from cost.** The worst position of the GFC, MA, fell 62% from its high while still showing a profit against its cost basis; a cost-based stop never fires on it.
- **Scaled by the name’s own volatility.** 2.5 monthly sigmas is about 35% for TEAM and 14% for PGR. A flat 10% stop – the intuitive choice – treats normal movement in a volatile name as a signal, and halves the return in simulation (27.5% to 13.5%); a flat 15% costs 7.6pp.
- **No re-entry until the trend repairs, and the proceeds stay in cash.** The optimiser ranks on five-year means, which keep pointing at the falling name: XYZ was 4% of the book when it first broke in 2021 and the optimiser rebuilt it to 28% on the way down. Redistributing the freed weight to the remaining names was far worse than holding cash (-45.8% against -27.8% in simulation), because the other names are usually in the same sell-off.

10.4 Result

The rule fires 76 times in 19 years, 3.9 times a year, at a median of 21.1% below the position's peak. The book stays largely invested – median invested share 88% – and de-risks in the episodes: through the 2022 drawdown it averages 64% invested with 7 stops, and through the GFC 86% with 9.

Arm, 30% spread	Return	Sharpe	Max DD	Top-5 DD	GFC	COVID	2022
S&P 500 (price)	10.18%	0.54	-55.2%	-28.0%	-55.2%	-33.7%	-13.3%
GCLPM baseline	30.59%	1.16	-45.1%	–	-15.6%	-28.4%	-31.3%
Corrected DCC	29.37%	1.12	-46.4%	–	-20.7%	-28.2%	-33.7%
+ state band + top-10	32.35%	1.26	-43.5%	-34.5%	+1.6%	-29.7%	-32.2%
+ stop, 21-day re-entry	31.40%	1.41	-27.9%	-24.4%	+14.7%	-27.3%	-16.0%
+ stop, 63-day re-entry	30.93%	1.45	-27.5%	-23.5%	+23.6%	-26.8%	-16.0%

The S&P 500 row is the benchmark the book is measured against: over the same window it compounds at 10.18% with a –55.2% drawdown, so the allocator's drawdown is already well inside the index's, and the stop brings it to half.

Against the arm it modifies, the stop costs 0.95pp of annualised return and buys 15.6pp of maximum drawdown, 10.1pp on the mean of the five deepest episodes and 0.15 of Sharpe. Against the GCLPM baseline the book earns +0.81pp more with a drawdown 17.2pp shallower and a Sharpe 0.25 higher; against the index it earns three times the return with half the drawdown.

The re-entry trigger was tested both ways. A 21-day moving average returns the name sooner and earns 0.47pp more than a 63-day one, at 0.4pp of maximum drawdown and 0.04 of Sharpe; the difference concentrates in 2008, where the slower trigger keeps the book out of the decline for longer (+23.6% against +14.7% through the GFC window). Both are well inside what a single path resolves. The 21-day trigger is the adopted setting.

The effect size was predicted before the backtest. The same rule applied as an overlay to the optimiser's targets forecast 75 stops, a drawdown near –28% and roughly flat return; the backtest produced 74 stops and –27.5%, with the return cost 1.4pp rather than zero, which is what execution and fees add.

10.5 The controls are not specific to DCC

The band and the stop act on the allocator's orders, not on its covariance, so they can be put on the lower-partial-moment baseline as readily as on DCC. That is worth testing rather than assuming, because the baseline has been ahead of the corrected estimator throughout this note, and because the top-10 reference set — the one piece that is DCC-specific, since GCLPM has no (α, β) to fit — has nothing to do on it.

Arm, 30% spread	Return	Sharpe	Max DD	Trades	Up capt.	Down capt.
GCLPM baseline	30.59%	1.16	-45.1%	710	155.2%	71.7%
+ state band	32.11%	1.24	-44.1%	643	153.4%	61.9%
+ state band + stop	29.82%	1.32	-28.5%	637	130.4%	44.5%
Corrected DCC + band + top-10	32.35%	1.26	-43.5%	733	153.8%	61.8%
+ stop	31.40%	1.43	-27.9%	714	132.7%	39.7%

Both allocators respond the same way. The band cuts downside capture from roughly 72% to 62% while leaving upside capture untouched, and adds about 1.5pp of return on fewer trades; on

the baseline it turns the GFC from -15.6% into $+4.7\%$. The stop then takes downside capture to 40–45% and halves the maximum drawdown, giving up some upside to do it. Neither control needs the DCC estimator to work.

Where the estimator still shows is the downside. With the full set of controls the corrected DCC book captures 39.7% of the index’s falls against 44.5% for the baseline, with a downside beta of 0.57 against 0.65, and that — not the return, which differs by 1.58pp — is what separates their Sharpe ratios (1.43 against 1.32). The gap is inside what a single path resolves, but it is the one difference between the two allocators that appears consistently at both control levels.

The practical reading is that the drawdown controls are the larger effect and the estimator choice the smaller one: moving from the baseline as it ran to either allocator with band and stop is worth roughly 0.16 of Sharpe and 17pp of maximum drawdown, while the choice between GCLPM and corrected DCC inside that is worth about 0.11 of Sharpe.

10.6 Re-entry variants that were rejected

Three other ways to return a stopped name to the book were measured on the optimiser’s targets, and all three cost more than they earned. Buying back the moment the ban lifts, rather than waiting for the next rebalance, loses 1.2pp of return and 10pp of drawdown: the rebalance calendar acts as a second confirmation, since the optimiser has to re-derive that it still wants the name. Parking the freed weight in the benchmark instead of cash earns 1.8pp but gives back 13pp of drawdown, because it restores the market exposure the stop had just removed — the GFC window degrades from -27.7% to -43.2% . Choosing the trigger by what the market was doing when the stop fired — fast when the benchmark was falling, slow when the name broke on its own — behaves as intended episode by episode but lands below the plain 21-day rule on both return and Sharpe.

The motivation for all three was that stopped names recover: 61% are above the stop price within 63 sessions, and the median move given away between the stop and the re-entry is $+14.7\%$. That figure is real, but it cannot be collected without also re-entering the 39% that keep falling.

10.7 Caveats

$k = 2.5$ was the best of three values tried on this path, though the 1.5/2.0/2.5 frontier is smooth and all three reduce the drawdown; the floor and cap were not tuned. Turnover rises by about a third in the overlay, and the backtest trades slightly *less* (694 against 733) because a stopped name is not repurchased for months. The rule does not help in a crash as fast as COVID, where 23 sessions separate the peak from the trough. It is off by default and is enabled with `stop_vol_k: 2.5` (`stop_reentry_ma` defaults to 21); adopting it in live trading is a separate decision, and it would need the stop orders to bypass the no-trade band as they do here.

11 The live universe, December 2025 to August 2026

The four arms were re-run on the universe the live configuration actually carries — the twenty names in `strategies.core.assets` as committed: URI, AAPL, AVGO, GOOGL, JPM, NVDA, CAMT, HWM, LLY, KLAC, ASML, CDNS, JNJ, SNDK, MU, LRCX, VLO, PANW, FTNT and GEV — with the live settings (25% position cap, 30-day horizon, 30-day rebalance, 5% trade band, 0.2% cash reserve, no momentum bias, the production 2% fee) at a 60% spread.

11.1 A spread has a feasibility ceiling

The allocator sets $\text{target} = \min(\mu_{\text{bench}} \cdot 252 + \text{spread}, \max_i \mu_i \cdot 252)$. Under a position cap c the most a long-only book can earn is the average mean of its best $1/c$ names, so the target is

reachable only while

$$\text{spread} \leq \bar{\mu}_{\text{top } 1/c} - \mu_{\text{bench}} \quad (14)$$

Above that the quadratic program has no feasible point. It does not degrade gracefully: the solver returns **unknown**, the allocator logs “Suboptimal solution”, and the book simply does not trade.

This is not hypothetical. On the sixteen-name universe in the working tree the ceiling over this window runs between 25.7% and 29.2%, so at a 60% spread five of six rebalances had no solution, the book sat in cash from December to July, and *all four arms returned identical numbers* — when the return constraint alone dictates the holdings, the covariance has no freedom to act on and the estimator, the band, the reference set and the stop cannot differ from one another.

On the committed twenty-name universe the ceiling is 42.6% in December and between 87% and 106% from January, because SNDK and GEV carry five-year means of 88% to 276%. A 60% spread is therefore feasible here from January, and the book begins trading on 2026-02-24. SNDK’s mean is measured over roughly eighteen months of history — it is the same short-history name that motivated the tiered estimation window of Section 3 — so the feasibility of this spread rests on one young estimate.

11.2 Result

Arm, 60% spread	Total	Annualised	Vol	Max DD	Trades	Stops
Baseline (no DCC)	34.74%	78.49%	26.8%	-12.0%	24	0
Corrected DCC	31.10%	69.22%	27.5%	-12.2%	21	0
+ state band + top-10	32.12%	71.80%	27.5%	-12.0%	22	0
+ volatility-scaled stop	30.26%	67.13%	26.4%	-9.8%	23	1

The S&P 500 returned 13.01% over the same 135 sessions. The ordering matches the long study in sign — the lower-partial-moment baseline ahead of the corrected DCC — but six months resolves nothing: the block bootstrap of Section 7 puts the resolution limit at roughly $\pm 3\text{pp}$ of *annualised* return over nineteen years.

Two observations survive the sample-size objection because they are structural rather than statistical. First, every arm holds JNJ at the 25% cap in every month, with SNDK, GEV and LLY supplying the return constraint: minimising $w'Hw$ subject to a demanding return floor puts the cap on the lowest-variance name that fits and buys the return elsewhere. Second, the stop fired once, on SNDK on 2026-07-16 at 39.6% below its peak, and cut the July drawdown from -12.0% to -9.8% for 1.9pp of return. SNDK fell a further 28% after the exit and then recovered to finish 11% *above* the stop price by 31 August: the rule caught the fall and missed the rebound, which is the cost side of Section 10 showing up in a window that contains no real drawdown.

12 What the estimator cannot fix

Sections 3 to 11 improved the inputs to the allocator: a correct DCC recursion, a horizon-consistent forecast, tiered estimation windows, a stable reference set. Each was worth one to two points of annualised return. The controls of Section 10 — a band and a stop, which touch no estimate at all — were worth ten times that on drawdown. That imbalance was the motivating question for this section: if better estimates buy so little, what is deciding the book?

12.1 Thesis

The allocator minimises $w'Hw$ subject to $\mu'w \geq \text{target}$, $w \geq 0$, $w \leq c$ and $\sum w = 1$. The covariance matrix only chooses *among* the books that clear the return floor. If that floor admits very few books, then no estimator change can matter, because there is nothing left to choose. The thesis is therefore that the binding object is the return constraint, not the risk model — and that this is measurable directly, without any backtest, by counting how much room the constraint leaves.

12.2 The floor is unreachable a third of the time

Under a position cap c the most a long-only book can earn is the cap-weighted mean of its best $\lceil 1/c \rceil$ names; call that the ceiling. Over the 154 rebalances of the 2003–2022 study, at a 30% spread and a 30% cap:

Measured over 154 rebalances	
Target pinned at $\max_i \mu_i$ rather than the spread	25 (16%)
Target <i>above</i> the feasibility ceiling	54 (35%)
Median slack (ceiling less target)	5.5pp
Slack at the 25th percentile	−2.8pp
Median number of names clearing the floor, of 26	2
Rebalances with fewer than four names clearing the floor	119 (77%)

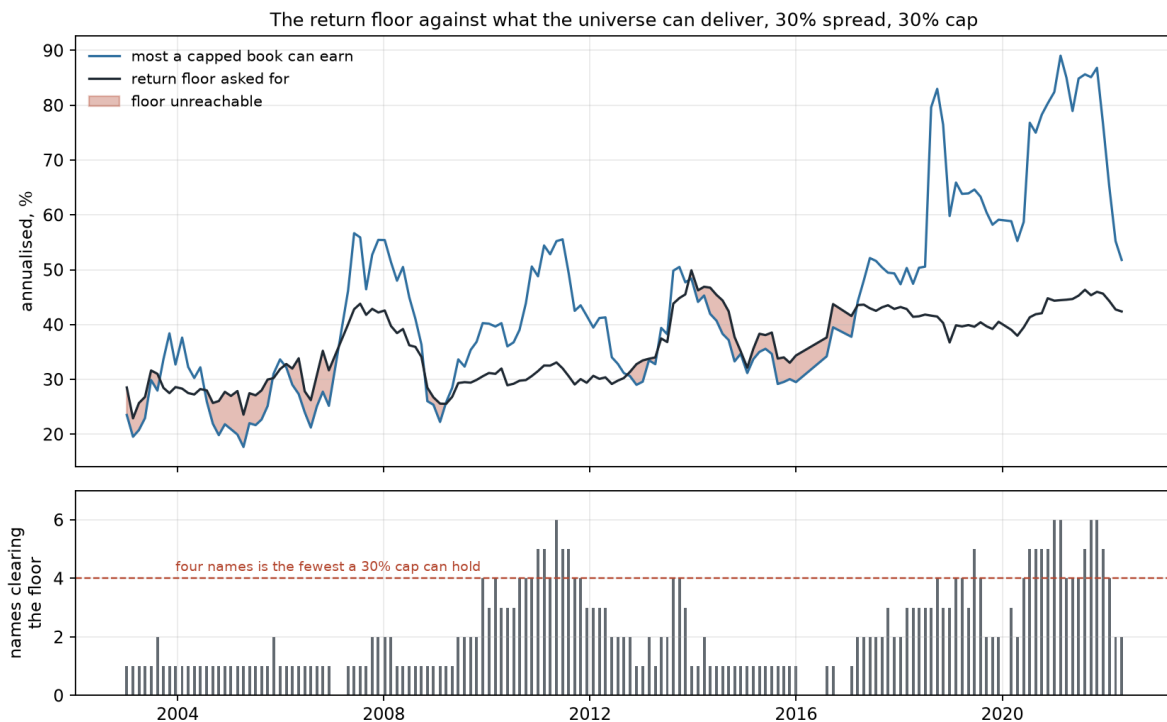


Figure 9: What is asked of the book against what the universe can give it. The shaded stretches are the 35% of rebalances where no capped long-only book can reach the floor; the panel below counts the names whose own mean return clears it, against the four a 30% cap must hold.

The median rebalance has two names in the whole universe whose own mean return reaches the floor. The book must hold them, at the cap, and make up the rest with whatever the

constraint will tolerate. This is the explanation for the null results of the screen in the next section, and it is not a modelling subtlety: it is arithmetic on the configuration.

12.3 When the floor is unreachable the book is a corner

On the 54 rebalances where the target exceeds the ceiling the programme is infeasible. It does not fail cleanly. Solving each date and recording what came back:

	Rebalances	Names held	At the cap	Monthly vol
Floor reachable	100	7.7	0.9	8.14%
Floor unreachable	54	4.0	3.0	10.16%

On every one of the 54 the realised portfolio mean equals the ceiling to four decimal places and the book is four names with three of them at the 30% cap: the solver returns the *maximum-return corner*, and reports `optimal` while doing it. A quarter more volatility than usual is taken on in exchange for a return the book cannot deliver, and nothing in the logs says so.

This refines the claim made in Section 11. There the same infeasibility produced no trades at all — the solver raised, the engine logged “Suboptimal solution”, and the book sat in cash. Both are the same defect seen from two sides: an infeasible return floor has no defined behaviour, so whether it yields a stale book or a maximally concentrated one depends on which way the interior-point method happens to fail. Six of the 154 dates raised; the other 54 went to the corner.

12.4 Spread and cap are the risk dial

The ceiling is a function of the cap, and the floor a function of the spread, so the two settings jointly decide how often the allocator is an optimiser at all. Re-solving all 154 rebalances over a grid, the share of dates on which the floor is unreachable:

Spread \ cap	20%	25%	30%	40%	50%
15%	0.6	0.0	0.0	0.0	0.0
20%	12.3	9.7	5.8	5.2	3.9
25%	30.5	26.6	20.8	13.6	11.0
30%	51.3	38.3	35.1	29.2	25.3
40%	76.6	74.7	67.5	53.2	46.1
60%	98.1	91.6	89.6	87.0	80.5

and what those books did, at a 30% cap:

Spread	Names held	Annualised	Sharpe	Max DD
15%	10.8	18.72%	1.12	-37.5%
20%	9.1	20.89%	1.06	-39.7%
25%	7.7	24.01%	1.04	-44.6%
30%	6.5	28.72%	1.09	-49.8%
40%	4.6	33.21%	1.05	-57.9%
60%	4.0	29.00%	1.02	-56.0%

The spread moves annualised return by ten points and maximum drawdown by twenty. Every estimator change in this paper moves return by one or two. The operator’s two configuration numbers dominate the modelling, and they do it through the same mechanism the corner

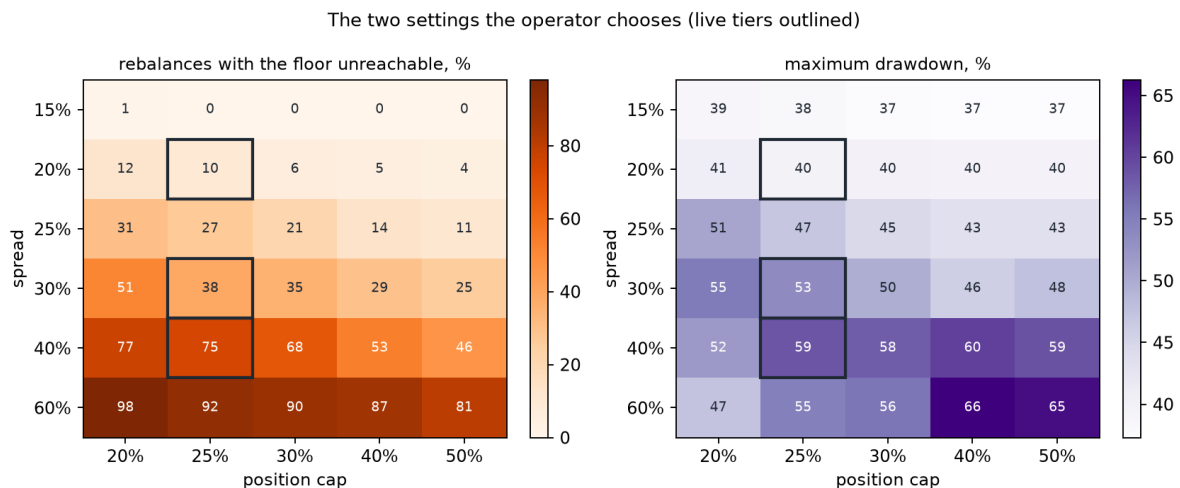


Figure 10: The share of rebalances on which the floor cannot be reached, and the maximum drawdown that results, over the two settings the operator chooses. The outlined cells are the three live tiers.

diagnostic exposes: raising the spread reduces the number of admissible books until only the highest-mean names remain.

Two further readings of the grid are worth recording. A *tighter* cap makes drawdown *worse* at high spreads — 51.3% of dates infeasible at a 20% cap against 25.3% at 50% — which is why the 20% cap tested in Section 10 lost: tightening the cap lowers the ceiling, which pushes more dates into the corner. And Sharpe is flat, between 1.02 and 1.12, across the whole spread range at a 30% cap: the floor buys return by taking risk at roughly a fair exchange rate, not by finding something the covariance missed.

12.5 What this implies for the live tiers

The committed live configuration runs three tiers at spreads of 40%, 30% and 20% with a 25% cap, and drops to a 15% `recessionary_spread` when the business-cycle predictor calls a recession. Read against the grid, on a universe of this size the aggressive tier has its floor unreachable on roughly three quarters of rebalances, the balanced tier on roughly a third, and the conservative tier on one date in ten. On the aggressive tier the allocator is therefore not a minimum-variance optimiser most of the time: it is a top-four selector by five-year mean return, held at the cap. That is a defensible product — it is what the return floor asks for — but it should be a choice rather than a side effect, and it means the estimator work in this paper has the least room to act exactly where the most risk is being taken.

There is a second, more mundane consequence. Rebuilding the study at the aggressive tier's settings — a 40% spread with a 25% cap — the optimiser fails outright on 42 of the 154 rebalances, against 6 at the settings used through this paper. A failed rebalance is not a neutral event: the previous book is carried forward, so the tier that asks for the most return is also the one that most often stops asking for anything at all. That is worth a monitor regardless of what else is decided, and it costs nothing to compute — the ceiling is four multiplications of the mean-return vector.

The remedies are not symmetrical. Lowering the spread to the reachable region restores the optimiser but gives up ten points of return. Keeping the spread and defending the concentration with position-level controls is what Section 10 does, and it is cheaper. The next two sections screen for something better than either.

13 A screen of allocator changes

13.1 Method

Fifty-nine variants were built on a screening harness rather than by backtest. The harness replays the optimiser at the same 154 rebalance dates on the same cached covariance forecasts and holds each target to the next date, so a difference between two arms is attributable to the change and nothing else. It reproduces the backtest’s own targets and, on the base configuration, its path: 28.72% annualised, 1.09 Sharpe, -49.8% maximum drawdown. Levels are not backtest levels — there is no band, no stop, no fee and no execution — but rankings are, and one arm takes seconds instead of ninety minutes.

Because the adopted configuration puts a band and a stop in front of the optimiser, every arm is also replayed through those: the per-name state band of Section 10, the volatility-scaled stop with 21-day re-entry, and 10bp of one-way cost. A change that only helps before the controls is not a candidate, because it is not what would be deployed.

Two limits are stated up front. A twelve-month block bootstrap of monthly differences gives a 95% interval of roughly $\pm 3\text{pp}$ on annualised return, so most of the differences below are not individually resolvable on one path; and the controls are path dependent, so a one-point weight difference can move a stop by a week and five points of drawdown with it. Conclusions are therefore drawn from agreement across independent implementations, from behaviour in identified episodes, and from mechanism — not from the ranking of a single number.

13.2 Theses and verdicts

Change	Thesis	Verdict
Feasible target	remove the silent infeasibility	enabler, null alone
Risk-return objective	drop the floor for a trade-off	real, wrong product
Shrink μ to the mean	stabilise the weakest input	confounded, then rejected
History gate on μ	a young name must not set the target	no effect here
Equilibrium (beta) prior	shrink with economic content	same as plain shrinkage
Nonlinear shrinkage	a better-conditioned matrix	null
Weight ensembling	average away near-degenerate caps	null
Mean-CDaR	optimise drawdown directly	rejected
Co-semivariance	GCLPM’s intent, done convexly	small, real
Downside-weighted DCC	close the crash-correlation gap	small, real
DCC volatilities, LPM	both parents at once	small, real
First-component cap	the drawdowns were factor bets	never bound
Benchmark beta cap	the same, measured properly	infeasible with the floor
Twelve-month momentum	choose the horizon deliberately	decisively worse
Step back from the ceiling	stop chasing an unreachable floor	fair-price trade
Turnover in the objective	price the trade, do not filter it	screened well, backtest rejected
Spread scaled by stress	ask for less when it is dangerous	screened well, backtest rejected

13.3 Results

Under the adopted controls, ordered by Sharpe. “top5 DD” is the mean of the five deepest non-overlapping drawdowns, which is less hostage to one episode than the maximum.

Arm	Annualised	Sharpe	Max DD	top5 DD	Turnover	Stops
Turnover penalty + stress spread	28.53%	1.35	-36.4%	-25.9%	185%	103
Turnover penalty, $c = 4$	29.74%	1.33	-43.4%	-28.3%	174%	105
Turnover penalty, $c = 2$	28.62%	1.29	-43.6%	-28.0%	179%	96
+ stress covariance	28.70%	1.28	-41.8%	-28.4%	185%	98
Both + stress covariance	26.65%	1.26	-35.3%	-26.8%	188%	105
Stress spread alone	25.52%	1.25	-33.5%	-24.1%	242%	83
Both + co-semivariance	26.51%	1.25	-30.1%	-26.0%	187%	95
Shrunk μ , capped target, penalty	32.81%	1.25	-43.3%	-31.7%	199%	72
Lower-partial-moment baseline	27.63%	1.25	-38.3%	-27.9%	203%	92
Downside-weighted DCC	27.73%	1.24	-41.3%	-28.3%	226%	82
Nonlinear shrinkage	27.40%	1.21	-36.3%	-28.1%	231%	78
Weight ensembling	26.91%	1.18	-41.4%	-29.2%	234%	79
Feasible target	26.99%	1.18	-41.4%	-29.3%	242%	79
Adopted configuration	26.68%	1.17	-41.4%	-29.2%	232%	79
Risk-return, $\lambda = 40$	13.59%	1.21	-32.5%	-16.4%	186%	121
Twelve-month momentum	13.75%	0.80	-40.7%	-26.1%	278%	94

13.4 Following the two algorithms' own intent

One arm was built to answer a different question: not which published method is best, but what the two allocators already in the codebase are each trying to say, and whether both can be said at once. GCLPM prices *downside* co-movement but its matrix is static. DCC is dynamic but symmetric, and Section 9 measured the consequence — it understates crash correlation by about 70%. The synthesis is to take the conditional volatilities from the DCC recursion and the dependence structure from the co-lower-partial-moment matrix, then project the product back to positive definiteness:

$$H_t = D_t R^- D_t, \quad R^- = \text{corr}(\Sigma^-), \quad \Sigma_{ij}^- = \frac{1}{T} \sum_t \min(r_{it} - \tau, 0) \min(r_{jt} - \tau, 0) \quad (15)$$

with τ the daily target return and D_t the conditional standard deviations. Blending R^- with the DCC correlation at equal weight is the half-way version.

It works, and it is small. Before the controls it is the best of the matrix family — 29.75% against 28.72%, with the tightest positive interval of any arm, $[-0.07, +2.10]$ pp — and it holds fewer names (5.7 against 6.5), which is the downside dependence doing its job. After the controls it is 28.51% at 1.22 Sharpe against 1.17, with the crisis window at -12.4% against -20.9% . But the plain co-semivariance matrix, with no DCC in it at all, lands in the same place, and so does the existing lower-partial-moment baseline. Three different implementations of “price the downside” agree on roughly $+0.05$ of Sharpe and two to three points of drawdown, which is the size of the effect — worth having, and an order of magnitude below the controls.

A practical note falls out of the same comparison. The live `_sclpm` matrix is built by a double Python loop of pairwise calls and took 431 seconds over the 154 rebalances; the co-semivariance matrix, a single matrix product, took 6, and the two behave indistinguishably here (29.43% at 1.12 against 29.47% at 1.12). If allocation latency ever matters, that is a seventy-fold saving available for no measured change in behaviour.

13.5 What was falsified, and why

- **Nonlinear shrinkage, weight ensembling and the feasible target are null before the controls** — 28.60%, 28.64% and 28.53% against the base 28.72%, with maximum

drawdown identical to the first decimal. This is the feasibility diagnostic showing through: when two names clear the floor, a better-conditioned matrix has nothing to do. Their apparent separation *after* the controls (nonlinear shrinkage at -36.3% against -41.4%) is stop timing, not estimation, and should not be read as an effect.

- **Shrinking the mean was confounded by the target rule.** Shrinking μ lowers $\max_i \mu_i$; the target $\min(\mu_{\text{bench}} + \text{spread}, \max_i \mu_i)$ then pins itself to a value only the single best name can reach, and the programme becomes infeasible. Shrinking by 25% lost 22 of 154 rebalances and by 50% lost 67, while concentrating the book from 6.5 names to 4.6. What looked like a return improvement (+5pp) was a concentration effect bought with 8pp of drawdown. Once the target is capped at the feasibility ceiling the arm solves and keeps most of the return, which is why the feasible target is recorded above as an enabler rather than as an improvement: it is null on its own and necessary for anything that touches μ .
- **The first-principal-component cap never bound.** An eigenvector of a 26-name covariance matrix has entries of order 0.2, so a cap of 0.45 on $\text{pc}'w$ was slack at every date — the 0.45 and 0.60 arms agree to the fourth decimal. Replacing it with the quantity actually intended, the book's benchmark beta, showed why it cannot be used as stated: capping beta at 0.8 makes the programme infeasible on 99 of 154 dates and at 0.6 on 126. The return floor *mandates* high beta; a beta cap and a demanding floor are contradictory instructions.
- **Mean-CDaR was worse on every axis** — 24.21% annualised, 0.92 Sharpe, -55.3% maximum drawdown, -33.4% through the financial crisis, and 58 of 154 dates unsolved. Optimising the deepest path of a 504-day scenario window is an in-sample exercise: the next drawdown has a different shape, and the linear programme has no shrinkage to protect it.
- **Twelve-month momentum, the classical horizon, is much worse here** — 13.75% against 28.72% before the controls, and worse again with them. The five-year mean is not a lazy choice; on this universe the longer horizon is the better estimator, and skipping the most recent month (the standard reversal correction) makes it worse still.
- **Stepping back from the ceiling is a fair-price trade, not a free fix.** Asking for 95% of the ceiling when the floor is unreachable — the direct remedy for the corner — gives up 1.3pp of annualised return for 5.5pp of maximum drawdown, at unchanged Sharpe. Deeper step-backs trade at the same rate. So the corner is a real defect in the sense that it is undefined behaviour, but in this sample the concentration it produces was paid for: the highest-mean names of 2003–2022 earned their volatility. The honest conclusion is that the corner should be made explicit rather than removed.

14 Two levers that the backtest rejected

Two changes came out of the screen ahead of everything else, and both were carried through to a full backtest. Neither survived it. The screen, the backtest, the reason to believe the backtest and the reason the screen was wrong are all recorded here, because the failure is more instructive than the candidates were.

14.1 Charging the objective for turnover

The no-trade band suppresses a trade *after* the optimiser has asked for it. It cannot distinguish a trade that buys a real reduction in risk from one that chases estimation noise, because by the

time the band sees the order the reasoning that produced it has been discarded. It is also blunt in the one place it matters most: in March 2020 the widened stress band locked in a book that wanted to change, and the episode return went from -28.3% without the band to -37.1% with it.

The alternative is to put the cost where the decision is made. Minimise

$$w'Hw + c \sum_i (w_i - w_{\text{prev},i})^2 \quad \text{subject to the same floor, caps and group limits} \quad (16)$$

which expands to $w'(H + cI)w - 2cw'_{\text{prev}}w$ and so costs one diagonal addition and one linear term — the optimiser already carries a zero linear term, so this is a two-line change. Now a trade happens when the risk it removes is worth more than the cost of making it, at every level of stress, with no state machine. The quadratic form was chosen over the absolute-value form deliberately: the ℓ_1 version needs $2n$ variables and two redundant constraint blocks, and with the feasible set as thin as Section 13 shows it to be, the interior-point solver failed on a third of the dates for reasons that had nothing to do with the economics. c is expressed in units of the average name's variance, so one setting travels across universes and horizons.

On the screen it looked like the best change in the study:

Screen, turnover penalty	Annualised	Sharpe	Max DD	top5 DD	Turnover
none (adopted configuration)	26.68%	1.17	-41.4%	-29.2%	232%
$c = 0.5$	27.87%	1.24	-43.9%	-28.8%	206%
$c = 1$	28.45%	1.26	-44.4%	-28.7%	192%
$c = 2$	28.62%	1.29	-43.6%	-28.0%	179%
$c = 4$	29.74%	1.33	-43.4%	-28.3%	174%
$c = 8$	29.62%	1.32	-41.0%	-27.2%	168%

Return and Sharpe rose monotonically and turnover fell monotonically over a sixteen-fold range of the parameter, which is normally what separates a lever from a fitted number.

14.2 Asking for less return when the tape is stressed

Section 13 showed that the return floor is the mechanism that concentrates the book, and that it concentrates hardest when cross-sectional dispersion is low. It is also completely blind to conditions: the same spread is demanded in a calm October and a falling March.

Production already contains the idea in discrete form. When the business-cycle predictor calls a recession, `recessionary_spread` drops the demand to 15%. The objection is the one already made against a drawdown-triggered band in Section 10: a recession call arrives after the fact. The same argument that chose trailing volatility as the band's signal applies here, so the spread is scaled continuously by the same measure,

$$\text{spread}_t = \text{spread} \cdot (f + (1 - f)(1 - s_t)), \quad s_t = \text{clip}\left(\frac{\sigma_t^{\text{bench}} - 0.20}{0.30 - 0.20}, 0, 1\right) \quad (17)$$

with f the floor the spread is allowed to fall to. On the screen this was the best drawdown lever found: at $f = 0.5$, maximum drawdown -33.5% against -41.4% , the top-five average -24.1% against -29.2% , and the pandemic episode — the one the band made worse — -15.5% against -35.9% , for about 1.2pp of annualised return. Combined with the turnover penalty the screen put the pair at 1.35 Sharpe against 1.17, with a flat optimum over $c \in [2, 4]$ and $f \in [0.5, 0.85]$, an edge that held across eight different stop and band settings and in both halves of the sample.

14.3 What the backtest said

Both levers were implemented in the backtester behind their own configuration keys and run over the same window, universe and settings as the adopted configuration — 2003-01-08 to 2022-05-31 at a 30% spread, with the corrected estimator, the top-ten reference set, the state band and the volatility stop all in place.

Backtest, 30% spread	Annualised	Sharpe	Max DD	Down capture	Trades	Turnover
Lower-partial-moment baseline	30.59%	1.16	-45.1%	71.7%	710	—
Corrected DCC	29.37%	1.12	-46.4%	73.2%	776	—
+ band + top-10 reference	32.35%	1.26	-43.5%	61.8%	733	—
+ stop (adopted)	31.40%	1.43	-27.9%	39.7%	714	366%
+ turnover penalty	30.41%	1.39	-27.5%	49.0%	686	249%
+ stress-scaled spread	27.13%	1.29	-28.7%	51.8%	813	460%
+ both	27.42%	1.33	-28.0%	54.4%	794	259%

The turnover penalty costs 0.99pp of return and 0.04 of Sharpe, and makes downside capture materially worse, 49.0% against 39.7%. The stress-scaled spread costs 4.27pp and 0.14. Neither improves the drawdown by more than a rounding error. The code for both was removed from both repositories; the adopted configuration of Section 10 stands unchanged.

14.4 Why the screen was wrong

The first explanation to reach for — that the penalty never bit, because the band was already suppressing the same trades — is wrong, and the transactions say so. Traded value over the window falls from \$308m to \$173m, a 44% reduction, and annualised turnover against the average book from 366% to 249%. The penalty did exactly what it was designed to do. The trades it removed were, on balance, trades worth making: the book adapted more slowly, and the downside capture figure is where that shows up.

That is the same conclusion Section 8 reached by a different route, now confirmed on the fixed engine with a proper implementation rather than rejected on a defective one. If turnover were itself the problem, less of it would be monotonically better. It is not. What the band buys is not *less* trading but *differently timed* trading — it blocks repositioning into the worst prices on the least reliable estimates, and it leaves calm-market rebalancing alone. A penalty in the objective cannot make that distinction, because it charges the same price for a trade in March 2020 as for one in a quiet October, which is precisely the state dependence the band exists to express.

The stress-scaled spread failed differently and more simply. Its annualised turnover rose from 366% to 460%: a target that moves every month with volatility rebuilds the book continuously, and it rebuilds it hardest in exactly the falling tape where the rebuild is most expensive. The screen showed the same signature — turnover up from 232% to 242% — and it was read as an acceptable price for the drawdown benefit. In the backtest the drawdown benefit did not appear and the cost did.

14.5 What this says about the harness

The screening harness reproduces the optimiser’s targets and, on the base configuration, its path. It ranked fifty-nine variants in the time one backtest takes, and the diagnostics it produced in Section 13 stand on their own. But it is a proxy for the *optimiser*, not for the deployed system, and these two arms mark out where the proxy fails:

- it models the per-name no-trade band as a single rule applied to a target vector, with no new-position bypass, no minimum trade size and no integer shares;
- it has no dividends, no taxes and no wash-sale disallowance, all of which interact with how often a position is closed;
- it charges a flat 10bp of one-way turnover, which flattens any change whose mechanism is “trade less” — and both of these were exactly that.

Which of those carries the difference is not established here, and it would take a separate experiment to find out. The usable rule is narrower and worth stating plainly: the harness may be trusted to rank changes to *what the optimiser holds*, and may not be trusted to price changes to *how often it trades*. Every conclusion in Section 14 that rests on holdings — the null estimator results, the infeasible floor, the beta-cap contradiction, the momentum horizon — is unaffected. The two that rested on trading frequency were the two that failed.

15 Where the return comes from

The work up to here was about risk: a correct estimator, a band, a stop. This section asks the other question — what is actually producing the return, and what is standing in its way — and answers it from the backtests’ own transaction tapes rather than from a screening harness. The book is taxable, so every comparison below carries an after-tax column, and that column changes one of this paper’s own conclusions.

15.1 The return is five names

Realised profit by name from the adopted run, marking open positions at their last traded price:

	Share of all profit
MKTX	17.2%
Top three (MKTX, NFLX, INMD)	46.1%
Top five (adding MSCI, FRHC)	68.8%
Names that lost money	9 of 33

Total profit was \$19.14m, and the nine losing names cost \$1.37m between them — 7% of the gross. Avoiding every loser perfectly would be worth about 0.35pp a year. The lever is the winners, and specifically how much of them the book is allowed to hold.

15.2 Both caps bind, nearly always

Constraint	Share of days it binds
ADR group cap, 10% across FRHC, SHOP and INMD	100% of the 1518 days those names are held
Per-name cap, 30%	79% of all days, 2.08 names at a time
BND group cap, 5%	93% of the days bonds are held

The three ADR names produced 29% of all profit from 22 trades and never once reached a 30% weight, because the group cap stopped them at 10% first. MKTX sat at the per-name cap on 1211 of the 2052 days it was held, NFLX on 1430 of 3482, BKNG on 997 of 1990. The optimiser has been asking for more of the winners for nineteen years and the configuration has been refusing.

15.3 The after-tax lens, which corrects Section 10

Arm	Pre-tax	After tax	Taxes	Max DD
Band + top-10 reference, no stop	32.35%	27.49%	\$10.0m	-43.5%
+ volatility stop (adopted)	31.40%	27.45%	\$7.4m	-27.9%

Section 10 reported that the stop costs about a point of return for its drawdown reduction. Pre-tax that is right: -0.95pp . After tax it costs 0.04pp , because the losses the stop realises are deductible and cut the tax bill by \$2.6m. For a taxable book the stop is very nearly free, and this paper has been understating the adopted configuration.

15.4 What the tax gap is, and what it is not

Gross 31.18% a year, 30.88% after fees, 27.46% after tax and fees. The 3.72pp gap is the largest single number in this study, and almost none of it is recoverable:

- **Lot selection is worth about 0.1pp.** Re-matching the existing tape under LIFO, highest-cost-first and an explicitly tax-minimising rule saves \$392k of \$5.55m. FIFO looks like the worst possible choice in a book that compounded 190-fold — the oldest lot is the cheapest — but the stop closes positions *entirely*, and when every lot is realised the order does not matter. Only partial sales have a choice.
- **Deferral was built, measured and removed.** Of the short-term gains, 22.6% sat on parcels held 300 to 365 days, and 89% of that was partial trims rather than stop exits: about twenty decisions in nineteen years, \$599k of addressable tax. A rule that holds such a trim until the lot turns long term duly cut taxes by \$0.8m — and returned 27.42% after tax against the adopted 27.45%. The trims it held cost exactly what the tax saved. The code was removed.
- **Two apparent opportunities are traps.** The engine's `naive_tax_loss_harvesting` flag harvests a loss and immediately repurchases, which disallows it under the wash-sale rule; and the 21-day stop re-entry is not a wash-sale problem, since only 1 of 339 sells was repurchased inside 30 days, or 13 counting the before-leg.

The honest reading is that the gap is the price of an active, concentrated, stop-protected strategy rather than an inefficiency waiting to be collected.

15.5 Relaxing the group cap

Arm	Pre-tax	After tax	Sharpe	Max DD	Down capture
Adopted, 30% spread and 30% cap	31.40%	27.45%	1.43	-27.9%	39.7%
+ ADR cap 25%	32.68%	28.68%	1.54	-28.4%	41.6%
+ ADR cap 50%	33.05%	29.59%	1.56	-28.4%	39.0%
+ ADR cap removed	33.07%	29.58%	1.56	-28.4%	39.0%
+ per-name cap 40%	31.60%	27.19%	1.39	-29.8%	44.7%
Live tier, 30% spread and 25% cap	26.50%	22.27%	1.28	-34.9%	55.8%
+ ADR cap 50%	30.26%	26.96%	1.52	-34.9%	46.1%

At the study settings the change is worth $+1.65\text{pp}$ pre-tax, $+2.14\text{pp}$ after tax and $+0.13$ of Sharpe, with maximum drawdown 0.5pp deeper and downside capture slightly better. At the live balanced tier it is worth $+3.76\text{pp}$, $+4.69\text{pp}$ after tax and $+0.24$ of Sharpe with the

drawdown unchanged. The curve flattens at 50%: removing the cap altogether gives 33.07% against 33.05%, so past that point the group cap is no longer what binds.

Relaxing the *per-name* cap does the opposite — it loses 0.26pp after tax, 0.04 of Sharpe and 1.9pp of drawdown — which is what stops this becoming a general rule about caps.

15.6 Is that conviction, or three lucky names?

The path-independent version of the question: across every rebalance and every name, do the positions the optimiser pushes to its cap outperform the ones it leaves below?

Forward horizon	At the cap	Below the cap	Difference
21 days	2.82%	1.85%	+0.98pp
63 days	7.72%	6.06%	+1.66pp
126 days	17.27%	13.23%	+4.04pp
63 days, 2003–2015	8.49%	4.11%	+4.38pp
63 days, 2016–2022	3.11%	8.74%	-5.63pp
63 days, excluding the ADR names	7.72%	5.24%	+2.48pp

Conviction at the per-name cap predicted well for thirteen years and inverted over the last six. That is not the same phenomenon as the ADR result, because those three names were pinned at 10% by the group cap and never came near the per-name one — and it is the reason the per-name relaxation fails while the group relaxation succeeds. The two caps constrain different things.

These observations overlap heavily, since one position sits at the cap across consecutive rebalances, so the effective sample is far smaller than the counts suggest and no significance is claimed. The weight-quartile version is not monotone and the correlation of target weight with forward 63-day return is 0.048. The mechanism is plausible; it is not a measured edge.

15.7 What was rejected

Change	Pre-tax	After tax	Sharpe
Adopted	31.40%	27.45%	1.43
Tax deferral, 65 days	31.37%	27.42%	1.41
Rebalance cadence 30 to 63 days	27.91%	24.23%	1.37
Per-name cap 30% to 20%	26.31%	21.65%	1.36
Spread 40% with a 25% cap	23.05%	19.12%	1.08

The last of these is worth a sentence, because it is Section 13 showing up in a live run rather than in a diagnostic: at a 40% spread the book placed 273 trades in nineteen years against the adopted 714. The floor is unreachable on three quarters of rebalances at that spread, so raising it does not concentrate the book — it stops the book trading.

15.8 Does it rest on one name?

FRHC is a Kazakh broker whose 2022 was idiosyncratic, and it is one of the five names carrying the book. Removing it from the universe entirely tests whether the case for the cap survives:

Arm	Pre-tax	After tax	Sharpe	Max DD	Down capture
Adopted, ADR 10%	31.40%	27.45%	1.43	-27.9%	39.7%
Adopted, ADR 50%	33.05%	29.59%	1.56	-28.4%	39.0%
FRHC removed, ADR 10%	29.19%	24.63%	1.29	-30.6%	53.9%
FRHC removed, ADR 50%	32.82%	28.85%	1.47	-30.1%	45.7%

It survives and strengthens. Without FRHC the relaxation is worth +4.22pp after tax against the +2.14pp it is worth with FRHC present, and it improves the drawdown and the downside capture rather than costing them. Dropping FRHC does cost the capped configuration 2.2pp, which confirms its place in the attribution — but the case for the cap does not rest on it.

15.9 Recommendation, and what it is not

Raise the ADR group cap from 10% to 50%. The after-tax gain is +2.14pp at the study settings, +4.22pp at those settings without FRHC, and +4.69pp at the live balanced tier. Sharpe improves at all three operating points, downside capture improves at all three, and maximum drawdown is unchanged or better at two of the three. The curve flattens at 50%.

This is leverage on the existing signal rather than a new source of return, and it should be argued for on that basis: a cap that binds on every single day a position is held is not expressing a risk preference, it is overriding the optimiser’s ranking. The symmetric point is that a future ADR name that goes wrong will cost proportionally more, and that the gain measured here comes entirely from 2016 onward, since the group was empty before SHOP entered in 2016. The cap presumably also exists for reasons this study cannot price — custody, liquidity, a concentration policy — and those are not backtest questions.

Nothing else in this round survived. The per-name cap at 40% loses; a tighter per-name cap at 20% loses heavily; a 63-day rebalance loses; a 40% spread stops the book trading; tax-lot selection is worth a tenth of a point; and the tax-deferral rule, which worked exactly as designed and cut the tax bill by \$0.8m, returned the saving straight back in foregone performance. The code for every rejected change has been removed from both repositories.

16 Interactive reports

The full backtests are published as interactive HTML reports (equity curves, allocation history, orders and statistics per run). The links are pre-signed and expire on 2026-09-28 at 23:24 UTC; regenerate them with `python aws/generate_report.py --run-uri <prefix>` from `xantos-backtest`, using the prefix listed in `study_results/dcc_study_s3_reports.json`.

- [Replication: GCLPM baseline, 9 spreads \(9 runs\)](#)
- [Replication: mgarch 0.3.0 as shipped, 9 spreads \(9 runs\)](#)
- [Replication: corrected DCC on the truncated window, 9 spreads \(9 runs\)](#)
- [Replication: corrected DCC with tiered windows, 9 spreads \(9 runs\)](#)
- [Crisis study: all 12 arms at a 30% spread \(12 runs\)](#)
- [Universe fix: state band with and without the top-10 reference set \(5 runs\)](#)
- [Band and stop on the GCLPM baseline, against the same controls on DCC \(7 runs\)](#)
- [Live universe \(committed, 20 names\) at a 60% spread, Dec 2025 to Aug 2026 \(5 runs\)](#)
- [Live universe, working-tree variant \(16 names\) at 17.5%, 25% and 60% \(12 runs\)](#)

17 Conclusions

The adopted configuration

Everything below is in the backtester. The estimator and the rebalance-trigger fix are unconditional; the state band is on by default; the reference set and the stop are opt-in. Adopting any of it in live trading is a separate decision.

- **Corrected DCC estimator** — `dcc: true`. A specification-faithful $\bar{Q}$ on standardised residuals, (α, β) actually fitted, the horizon scaled linearly in time, and tiered per-pair estimation windows. Unconditional: it is what `dcc.py` now does in both repositories.
- **Rebalance-trigger fix** — unconditional. The book acts on every target change rather than the two thirds that survived the old floating-point test.
- **State-dependent no-trade band** — `band_stress_enabled: true`, on by default. The band runs from `min_offset_for_trade_percent` up to `band_stress: 0.25` as benchmark volatility rises from `band_vol_calm: 0.20` to `band_vol_stress: 0.30`, on `band_signal: vol`.
- **Top-10 reference set** — `dcc_reference_top_n: 10`, opt-in. (α, β) are fitted on the ten highest mean returns plus the benchmark, so a name outside that set cannot move everyone else's correlations.
- **Volatility-scaled trailing stop** — `stop_vol_k: 2.5`, opt-in. Exit at 2.5 monthly sigmas below the position's high-water mark, bounded by `stop_floor: 0.08` and `stop_cap: 0.35`, proceeds to cash, re-entry once the price is back above its `stop_reentry_ma: 21`-day average with a minimum wait of `stop_min_wait: 21` sessions.

Together, against the GCLPM baseline over 2003–2022 at a 30% spread: return 30.59% to 31.40%, Sharpe 1.16 to 1.43, maximum drawdown -45.1% to -27.9% , downside capture 71.7% to 39.7%. Against the S&P 500 over the same window, three times the return with half the drawdown.

- **The estimator was defective, and the 2022 edge came from the defect.** Corrected to specification, DCC's return advantage at a 30% spread falls from +3.48pp to -1.22 pp once both the estimator and the backtester's rebalance trigger are fixed.
- **A better forecast does not fix crisis behaviour.** Model-side changes run over the full window moved return by up to 1.7pp but never improved the drawdown or the GFC.
- **When the book trades matters more than how much.** Blocking trades in stress is where the value is, and a constant band captures it only at one knife-edge setting. A quadratic turnover penalty was rejected here once on the defective engine and again, properly implemented, on the fixed one: it cut traded value by 44% and still lost 0.99pp of return and nine points of downside capture (Section 15). What the band buys is differently timed trading, not less of it.
- **The state-dependent band is the adopted change.** On the re-based runs the band and the reference set together add +1.76pp of return, +0.10 of Sharpe and a 1.6pp shallower drawdown than the baseline; the pre-fix measurement of the band alone was larger (+2.35pp), which is one reason Section 7 matters. It was chosen over a constant band because every state-band variant tried beats the baseline on all three goals while only one constant setting does, and it underperforms in slow, moderate-volatility declines such as 2022.

- **The controls are worth more than the estimator choice.** The band and the stop work on the baseline as well as on DCC — together 0.16 of Sharpe and 17pp of maximum drawdown — while the choice between the two allocators inside that is worth about 0.11 of Sharpe, and shows up in downside capture rather than return.
- **The drawdown is a position-level problem.** Portfolio-level constraints did not fix it — a tighter position cap made it worse — while a volatility-scaled trailing stop with a trend-based re-entry halved it. A flat 10% stop, the intuitive version, halves the return instead.
- **Zero weight was not zero influence.** The universe dependence came from the jointly fitted α ; fitting the dynamics on the top ten names removes it for laggards exactly, and by design not for winners. Over the full backtest it costs nothing, and where the added names were irrelevant it raised the share of days on which the book was unchanged from 81% to 92%.
- **The return floor, not the risk model, decides the book.** At a 30% spread the median rebalance has two names in a universe of twenty-six whose own mean return reaches the floor, and on 35% of rebalances no capped book can reach it at all. On those dates the optimiser returns a maximum-return corner — four names, three at the cap, a quarter more volatility — and reports success. This is why nonlinear shrinkage, weight ensembling, factor caps and a feasible target are all null to two decimal places: there is nothing left for the covariance to choose among.
- **The spread and the cap are the risk dial.** Across the grid, the spread moves annualised return by ten points and maximum drawdown by twenty, at a Sharpe that barely changes; every estimator change in this paper moves return by one or two. The aggressive live tier has its floor unreachable on roughly three quarters of rebalances, which makes it a top-four momentum selector rather than a minimum-variance optimiser. That should be a choice, not a side effect.
- **Two levers led a fifty-nine-arm screen and both failed the backtest.** A turnover penalty in the objective and a spread scaled down by benchmark volatility were worth +0.12 and +0.08 of Sharpe on the screen and cost 0.04 and 0.14 in the backtest. Their code was removed. The screen ranks what the optimiser *holds* reliably and what it *trades* unreliably, and both of these were changes to trading (Section 15).
- **The binding group cap was the return lever.** Raising the ADR cap from 10% to 50% is worth +2.14pp of after-tax return at the study settings, +4.22pp with the most idiosyncratic of its three names removed from the universe, and +4.69pp at the live balanced tier — with Sharpe and downside capture better at every operating point and the drawdown essentially unchanged. That cap bound on 100% of the days those names were held. Relaxing the *per-name* cap loses: the two constrain different things (Section 16).
- **For a taxable book the stop is nearly free.** It costs 0.95pp of return pre-tax and 0.04pp after tax, because the losses it realises are deductible and cut the tax bill by \$2.6m. Section 10 understated it.
- **The 3.72pp tax gap is mostly structural.** Lot selection is worth about 0.1pp, and a deferral rule that cut the tax bill by \$0.8m gave exactly that back in foregone performance. The gap is the price of an active, concentrated, stop-protected strategy rather than an inefficiency waiting to be collected.

- **What did not work is worth as much as what did.** Mean-CDaR, a benchmark-beta cap, twelve-month momentum, nonlinear shrinkage, weight ensembling, stepping back from the feasibility ceiling, a turnover penalty and a stress-scaled spread were all measured and rejected, most of them for structural reasons rather than for want of tuning. The adopted configuration is unchanged by this round of work.

18 Reproducing this

```
# unit tests for the estimator (xantos-ra)
pytest tests/test_dcc_covariance.py -q

# the four-arm replication (xantos-backtest)
XANTOS_MANAGEMENT_FEE_PCT=0.01 python run_backtest.py --config
    dcc_study_full_dcc_fixed.yml
XANTOS_MANAGEMENT_FEE_PCT=0.01 python run_backtest.py --config
    dcc_study_full_gclpm.yml
# arm B runs from a worktree pinned to the pre-fix commit
XANTOS_MANAGEMENT_FEE_PCT=0.01 python run_backtest.py --config
    dcc_study_full_mgarch.yml

# every study's configuration is stored next to its results
ls study_results/dcc_study_v*/config.yml

# crisis study and universe fix, 30 percent spread (the state band is now the
    default)
XANTOS_MANAGEMENT_FEE_PCT=0.01 python run_backtest.py --config study_results/
    dcc_study_v8_stateband/config.yml
XANTOS_MANAGEMENT_FEE_PCT=0.01 python run_backtest.py --config study_results/
    dcc_study_v9_reference/config.yml

# figures and tables for sections 7 and 8, then the PDF (xantos-ra)
python docs/analyze_crisis_study.py
python docs/build_writeup.py
```

References

- Engle, R. (2002). Dynamic Conditional Correlation — A Simple Class of Multivariate GARCH Models. *Journal of Business & Economic Statistics*, 20(3), 339–350.
- Engle, R. and Sheppard, K. (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Paper 8554.
- Xantos Labs LLC (2022). An Enhancement to Xantos' Systematic Model.